



# **Defining the metabolic phenotype of epithelial-mesenchymal transition in breast epithelium**

**Qiong Wang**

**Thesis for the degree of Philosophiae Doctor**

**Supervisor/s:**

Prof. Dr. Óttar Rolfsson

**Doctoral committee:**

Dr. Skarphéðinn Halldórsson

Prof. Margrét Þorsteinsdóttir

Dr. Sævar Ingþórsson

May 2022



**UNIVERSITY OF ICELAND**  
**SCHOOL OF HEALTH SCIENCES**

FACULTY OF MEDICINE



# **Auðkenning efnaskiptabreytinga samfara EMT í brjóstaþekjuvef**

**Qiong Wang**

**Ritgerð til doktorsgráðu**

**Leiðbeinandi/leiðbeinendur:**

Prof. Dr. Óttar Rolfsson

**Doktorsnefnd:**

Dr. Skarphéðinn Halldórsson

Prof. Margrét Þorsteinsdóttir

Dr. Sævar Ingþórsson

Maí 2022



**UNIVERSITY OF ICELAND**  
**SCHOOL OF HEALTH SCIENCES**

FACULTY OF MEDICINE

Thesis for a doctoral degree at the University of Iceland. All right reserved.  
No part of this publication may be reproduced in any form without the prior  
permission of the copyright holder.

© Qiong Wang 2022

ISBN (978-9935-9657-8-3)

Printing by Háskólaprent.

Reykjavik, Iceland 2022



## Ágrip

Brjóstakrabbamein er ein helsta orsök krabbameinsdauða meðal kvenna á Íslandi og um heim allan. Ein forsenda meinvarpsmyndunnar er talin vera sú að æxlisfrumur virkji þroskunarferli sem kallast bandvefsumbreyting (EMT). Bandvefsumbreyting (EMT) er frumuþroskunarferli þar sem þekjufrumur umbreytast í bandvefsfrumur með minni viðloðun og aukna skriðgetu. Talið er að æxlisfrumur geti virkjað EMT sem auki þá dreifingu æxlisfrumna milli vefja. EMT einkennist þó af fjölda frumusvipgerða (EMP) sem hafa mismunandi getu til meinvarpsmyndunnar. Efnaskiptabreytingar sem eiga sér stað samfara EMT og EMP eru illskilgreindar. Þetta grunnrannsóknarverkefni miðaði að því að einkenna efnaskiptasvipgerðir EMT og EMP í þeim tilgangi að auka skilning á framgangi brjóstakrabbameins, stuðla að framtíðar lyfjapróun og greina hugsanleg lífmörk fyrir illkynja æxli í brjóstum.

Proteingreining á EMT og EMP frumulíkönum leiddu í ljós breytingar í efnaskiptaferlum sykrunga. Ensímin UDP-glucose dehydrogenase (UGDH) og glutamine-fructose-6-phosphate transaminase 2 (GFPT2) sýndu mikinn breytileika í tjáningu samfara EMT annarsvegar og EMP hinsvegar. UGDH breytir UDP-glúkósa í UDP-glúkúrónsýru og tekur þátt í myndun hýalúrónans í utanfrumuefni. GFPT2 er hraðatakmarkandi ensímið í nýmyndun sykrunga og hvatar myndun UDP-*N*-asetýlglúkósamíns (UDP-GlcNAc) og O-GlycNAcylation proteina. Fylgni var á milli tjáningar UGDH og lífslíkum sjúklunga. Bæling á tjáningu UGDH hafði áhrif á frumufjölgun, frumuinnrás, og tjáningu EMT merksins SNAI1 og myndun glycerophosphocholines og *n*-acetylaspartats. Tjáning GFPT2 hafði einnig áhrif á frumufjölgun, frumuinnrás og stjórnun á EMT merkinu vimentin. Tjáning á GFPT2 sýndi fylgni við illkynja æxli í brjóstum með claudin-lág einkenni sérstaklega. GFPT2 tjáning eykst við oxunarág og bæling á GFPT2 hafði áhrif á innanfrumumagn cystathionines og á hvatberaensímið SQOR. Tjáningu á GFPT2 var stýrt af EGF og insúlíni boðskiptaferli, hugsanlega í gegnum GSK-3β. Að lokum voru boðskiptaferli skoðuð sérstaklega með skimun á fosforyleringu próteina. Breytileiki í fosforyleringu proteina samfara EMT og EMP var mestur vegna mismunandi virkni PDHKS og PAK1 prótein kínasa. Á heildina litið hefur verkefnið sýnt fram á að UGDH og GFPT2 eru veigamikil í efnaskiptum sykra samfara EMT og í EMP, í framgangi brjóstakrabbameins og að þessar breytingar eru mismunandi eftir undirflokkum brjóstakrabbameins.

**Lykilorð:**

Illkynja æxli í brjósti; Bandvefsumbreyting; GFPT2, UGDH, próteinefnagreining

## Abstract

Breast cancer is the leading cause of cancer deaths among women in Iceland and worldwide. Epithelial-mesenchymal transition (EMT) is a cellular developmental process where epithelial cells assume mesenchymal-like phenotypes through degradation of the extracellular matrix, loss of adhesions, and increased mobility. It is believed that dissemination of cancer cells occurs partly following EMT. EMT contains a spectrum of epithelial-mesenchymal intermediate cell states that impart different degrees of malignancy. The ability of cells to assume these states is termed epithelial-mesenchymal plasticity (EMP). This project aimed at characterizing the metabolic phenotypes of EMT and EMP to obtain knowledge of breast cancer progression and to identify biomarkers and potential therapeutic targets for breast cancer treatment.

Proteomics analysis of cell models of EMT and EMP revealed changes to enzymes involved in glycan metabolism. UDP-glucose dehydrogenase (UGDH) and glutamine-fructose-6-phosphate transaminase 2 (GFPT2) were identified as the topmost altered glycan metabolic enzymes in EMT and EMP, respectively. UGDH converts UDP-glucose into UDP-glucuronic acid and is involved in the formation of hyaluronan in the extracellular matrix. GFPT2 influences the downstream formation of UDP-*N*-acetylglucosamine (UDP-GlcNAc) and protein *O*-GlycNAcylation. UGDH was associated with patient survival, affected cell proliferation, cell invasion, and the expression of the EMT marker SNAI1. siRNA-mediated knockdown of UGDH influenced glycerophosphocholine (GPC) and increased *N*-acetylaspartate (NAA) levels. GFPT2 similarly influenced cell proliferation, migration, invasion, and expression of the EMT marker vimentin and was associated with claudin-low breast cancer. GFPT2 was shown to be a marker of oxidative stress, and knockdown of GFPT2 affected cystathionine levels and the mitochondrial enzyme sulfide quinone oxidoreductase (SQOR). Phosphoproteomics analysis indicated distinct phosphorylation profiles of epithelial *versus* mesenchymal cells. Specifically, pyruvate dehydrogenase kinases (PDHKs), serine/threonine-protein kinase (PAK1), and protein kinase A catalytic subunit  $\alpha$  (PKACA) were differentially regulated across the mesenchymal cell lines tested.

In conclusion, these results suggest that UGDH and GFPT2 are central to changes that occur within glycan metabolism following EMT and EMP, respectively. Both enzymes were associated with cancer progression and GFPT2, specifically, may serve as a biomarker for cellular oxidative stress and claudin-low breast cancer. The work furthermore implicates UGDH in GPC and NAA metabolism.

**Keywords:**

Breast cancer, EMT, GFPT2, UGDH, proteomics

## Acknowledgements

This doctoral study would not have been successfully conducted without the help of others. I want to take this opportunity to thank my supervisor, instructors and advisors, financial support, family and friends, and the institutions for their help.

Prof. Óttar Rolfsson has been a great supervisor and mentor for the last five years, and I felt lucky to work for him. I appreciate him always having a good temper, being positive, and being patient with me. He gave me helpful tips and guidance when I was confused with the direction of the study and helped me with all resources needed so that the project could continue smoothly. He was so dedicated and patient in helping me write and revise the presentations/proposals/manuscripts and made so much sense out of the confusion. He is a brilliant supervisor with skillful supervision who cares about students' wellbeing, so I could fully focus on my study with no additive external pressure.

I greatly appreciate the Icelandic Center for Research (RANNÍS), Göngum Saman, and the doctoral grant from the University of Iceland (Doktorsstyrkir) for financing my study. This project would not have been possible without their support.

Furthermore, I want to thank my Ph.D. committee members for their contributions and efforts: Dr. Skarphéðinn Halldórsson, Prof. Margrét Þorsteinsdóttir, and Dr. Sævar Ingþórsson, with special thanks to Dr. Skarphéðinn Halldórsson. He was very inspiring to me regarding his scientific mind and personality when I started my Ph.D. study.

In addition, my sincere thanks also go to every group member of the Center for Systems Biology lab. I am grateful to Sigurður Trausti Karvelsson for helping me with the computational work. I also want to express my thanks to people from other groups in the office for the help that I needed in the lab.

I would also like to recognize the invaluable help from the Stem Cell Research Unit. They were very supportive when I was in need of any experimental materials. Especially, I want to thank Sarah Sophie Steinhäuser. Sophie was super patient with me when I constantly asked her for help. She helped me with the cell growth factors and medium etc. that I had to borrow from her. She helped me with the datasets of the D492 cell

lines. She also helped me with all the questions that I had regarding these cell lines. Off work, she is an awesome friend and has left me with lots of happy memories of us together during that stressful period.

I am also grateful to Prof. Geir Slupphaug and the PROMEC lab at NTNU for taking me as a visiting student. I want to thank the contributions of Douglas Lamont and Amy Tavendale at the University of Dundee.

In the end, I would like to give my heartfelt thanks to my family. Both my Chinese family and my Icelandic family have been very supportive. I must express my deepest gratitude to my husband Freyr Jóhannsson. He has been so patient with me and encouraged me when I had setbacks. He endured my mood swings and constantly cheered me up. Not only has he been a dear friend by my side, but he is a great advisor with a brilliant mind and solid knowledge. He helped me analyze the metabolomic data, pointed out interesting findings, suggested plausible following experiments, and instructed my metabolomic experiments.

# Contents

<b>Ágrip</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>v</b>
<b>Acknowledgements</b> .....	<b>vii</b>
<b>Contents</b> .....	<b>ix</b>
<b>List of abbreviations</b> .....	<b>xiii</b>
<b>List of figures</b> .....	<b>xix</b>
<b>List of tables</b> .....	<b>xxi</b>
<b>List of original papers</b> .....	<b>xxiii</b>
<b>Declaration of contribution</b> .....	<b>xxiv</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Breast cancer .....	1
1.1.1 Breast cancer subtypes.....	1
1.1.2 Claudin-low breast cancer.....	2
1.1.3 Breast cancer metastasis .....	2
1.2 EMT.....	3
1.2.1 The definition of EMT .....	3
1.2.2 The history of EMT .....	4
1.2.3 EMT in cancer .....	5
1.2.4 The therapeutic implications of EMT .....	11
1.3 Cancer metabolism .....	12
1.3.1 Cancer and breast cancer metabolism .....	13
1.3.2 Metabolic reprogramming in metastasis and EMT.....	14
1.3.3 Crosstalk between oncogenic signaling and metabolism.....	17
1.3.4 Metabolic reprogramming in glycan metabolism.....	17
1.3.5 Metabolic reprogramming of one-carbon metabolism and the transsulfuration pathway .....	23
1.3.6 Metabolic reprogramming in ROS regulation and mitochondrial metabolism .....	24
1.3.7 GPC and NAA .....	25
1.4 The D492, HMLE, and PMC42 EMT cell models .....	27
1.4.1 The isogenic EMT cell lines D492, D492M, and D492HER2.....	28
1.4.2 The common EMT cell model HMLE-HMLEM .....	29
1.4.3 The MET cell model PMC42LA-PMC42ET .....	30

1.5	The multi-omics era in molecular biology.....	30
1.5.1	The rise of proteomics.....	31
1.5.2	Metabolomics: A missing puzzle in multi-omics.....	35
1.6	Our previous work on the metabolism of EMT.....	37
1.7	Project summary.....	38
<b>2</b>	<b>Aims.....</b>	<b>41</b>
<b>3</b>	<b>Materials and methods.....</b>	<b>43</b>
3.1	Cell culture.....	43
3.2	LFQ proteomics.....	44
3.2.1	Protein sample preparation.....	44
3.2.2	Peptide sample preparation.....	44
3.2.3	LC-MS/MS analysis.....	45
3.2.4	Peptide and protein identification and quantification.....	46
3.3	SILAC (phospho)proteomics.....	48
3.3.1	Protein sample preparation.....	48
3.3.2	Peptide sample preparation.....	48
3.3.3	LC-MS/MS analysis.....	50
3.3.4	Peptide and protein identification and quantification.....	50
3.4	Data Availability.....	51
3.5	Transient knockdown with siRNAs and quantitative reverse transcription PCR (RT-qPCR).....	52
3.6	Crystal violet assay.....	53
3.7	Metabolomics analysis.....	53
3.8	Untargeted metabolomics analysis.....	55
3.9	Cell proliferation assay.....	56
3.10	Scratch migration assay.....	56
3.11	Transwell invasion assay.....	57
3.12	Western blot.....	57
3.13	Hydrogen peroxide (H <sub>2</sub> O <sub>2</sub> ) and reduced glutathione (GSH) treatment and growth factor deprivation.....	58
3.14	Glutathione assay.....	58
3.15	Statistical analysis and bioinformatics.....	58
<b>4</b>	<b>Results and discussion.....</b>	<b>61</b>
4.1	Proteomic profiling of the EMT and EMP programs.....	61
4.1.1	Characterize the D492, HMLE, and PMC42 EMT cell models ( <i>paper I</i> ).....	61
4.1.2	Characterize the D492, D492M, and D492HER2 EMP cell model ( <i>paper II</i> ).....	67
4.2	Search for breast EMT markers.....	75



4.2.1	Common EMT markers were identified ( <i>paper I</i> ) .....	75
4.2.2	Distinct proteomic signatures of the EMP cell line ( <i>paper II</i> ).....	77
4.3	Metabolic alteration in breast EMT .....	82
4.3.1	The commonly altered metabolic targets post-EMT ( <i>paper I</i> ).....	82
4.3.2	Different metabolic profiles of the EMP cell model ( <i>paper II</i> ).....	84
4.4	Glycan metabolism alters in breast EMT .....	87
4.4.1	UGDH is highly expressed in mesenchymal cells ( <i>paper I</i> ).....	87
4.4.2	GFPT2 is associated with cancer malignancy ( <i>paper II</i> )	98
4.5	Reconstruction and upgrade of the D492 GEMs ( <i>paper III</i> ) .....	110
4.6	Deep mining of the phosphoproteomic data ( <i>unpublished data</i> ) .....	112
4.6.1	Phosphorylated metabolic alteration in EMT.....	113
4.6.2	Phosphorylated metabolic alteration in EMP .....	115
4.6.3	Phosphorylated metabolic alteration in cancer malignancy .....	117
4.6.4	Kinases responsible for the identified phosphosites .....	119
<b>5</b>	<b>Summary and conclusions.....</b>	<b>121</b>
	<b>References .....</b>	<b>125</b>
	<b>Paper I.....</b>	<b>179</b>
	<b>Paper II.....</b>	<b>207</b>
	<b>Paper III.....</b>	<b>235</b>
	<b>Appendix .....</b>	<b>247</b>



## List of abbreviations

2D	two-dimensional
2-HG	2-hydroxyglutarate
4-MU	4-Methylumbelliferone
ACLY	ATP citrate lyase
ACN	acetonitrile
ACP	acyl carrier protein
AGC	automatic gain control
AKR1B1	aldo-keto reductase 1 member B1
AKT	AKT serine/threonine protein kinase
ALDH1A3	dehydrogenase family 1 member A3
ALDOA	fructose-bisphosphate aldolase A
ALDOC	fructose-bisphosphate aldolase C
ALK5	transforming growth factor beta receptor I (TGFB1)
AML	acute myeloid leukemia
AMPK	AMP-activated protein kinase
ASL	argininosuccinate lyase
ASPA	aspartocylase
ATP	adenosine triphosphate
BMP-2	bone morphogenetic protein 2
BP	biological process
BRCA1/2	breast cancer gene 1/2
BRENCs	breast endothelial cells
CAFs	cancer-associated fibroblasts
CBS	cystathionine- $\beta$ -synthase
CACLE	the Cancer Cell Line Encyclopedia
CD104	integrin, beta 4 (ITGB4)
CD24	CD24 antigen, cluster of differentiation 24
CD34	hematopoietic progenitor cell antigen CD34
CD44	CD44 antigen, cluster of differentiation 44
CD70	cluster of differentiation 70
CDH1	<i>E</i> -cadherin
CDH2	<i>N</i> -cadherin
CHK $\alpha$	choline kinase alpha
CID	collision-induced dissociation
c-Krox	zinc finger and BTB domain-containing protein 7B
CNS	central nervous system
CoQ	coenzyme Q
CpG	DNA regions, CpG sites
CS/DS	chondroitin sulfate/dermatan sulfate
CSC	cancer stem cell
CTCs	circulating tumor cells
CV	coefficient of variation
DAG	diacylglycerols
DDA	data-dependent acquisition
DHCR7	7-dehydrocholesterol reductase
DIA	data-independent acquisition

DPYD	dihydropyrimidine dehydrogenase
DTT	dithiothreitol
E2A	transcription factor 3
E6/E7	HPV E6/7
ECM	extracellular matrix
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EMP	epithelial-mesenchymal plasticity
EMT	epithelial-mesenchymal transition
EMT-TFs	EMT-associated transcription factors
ERBB	receptor tyrosine-protein kinase
ERK	extracellular signal-regulated kinase
ESA	epithelial cell adhesion molecule (EpCAM)
ESI	electrospray ionization
ESR	estrogen receptor
EST	expressed-sequence tag
ETC	electron transport chain
F1,6BP	fructose 1,6 bisphosphate
FA	formic acid
FABP5	fatty acid binding protein
FAK	focal adhesion kinase
FASN	fatty acid synthase
FASP	filter-aided sample preparation
FBS	fetal bovine serum
FDFT1	farnesyl-diphosphate farnesyltransferase 1
FDR	false-discovery-rate
FGF	fibroblast growth factor
FH	fumarate hydratase
FN1	fibronectin
FOXO-1	forkhead box protein O1
FTMS	fourier transform mass spectrometry
FWHM	full width at half maximum
GAGs	glycosaminoglycans
GDPD5	glycerophosphodiester phosphodiesterase domain Containing 5
GEMMs	genetically engineered mouse models
GEMs	genome-scale metabolic network reconstructions
GES	gene expression studies
GFPT2	glutamine-fructose-6-phosphate transaminase 2
Glc	glucose
Gln	glutamine
GLO1	glyoxalase 1
glog	generalized logarithmic transformation
GLS1	glutaminase 1
GLUT1	glucose transporter
GO	gene ontology
GPC	glycerophosphocholine
GPC-PD	glycerophosphocholine phosphodiesterases
GSH	reduced glutathione
GSK-3 $\beta$	glycogen synthase kinase 3 beta
GSSG	oxidized glutathione
H <sub>2</sub> O <sub>2</sub>	hydrogen peroxide

H <sub>2</sub> S	hydrogen sulfide
HA	hyaluronan
HACD3	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3
HBP	hexosamine biosynthetic pathway
HCD	higher-energy C-trap dissociation
HER2	receptor tyrosine-protein kinase ERBB2
HGF	hepatocyte growth factor
HIF-1 $\alpha$	hypoxia-inducible factor-1 $\alpha$
HILIC	hydrophilic interaction liquid chromatography
HK	hexokinase
HMDB	human metabolome database
HMS	the Harvard Medical School
HPDL	4-hydroxyphenylpyruvate dioxygenase-like protein
HPP	the human proteome project
HPV	human papilloma virus
HS	heparin/heparan sulfate
hTERT	telomerase reverse transcriptase in humans
IAA	iodoacetamide
iBAQ	intensity-based absolute quantification
IDH	isocitrate dehydrogenase
IGF	insulin-like growth factor
IPA	Ingenuity Pathway Analysis
IRE1 $\alpha$	inositol-requiring enzyme
ITMS	ion trap mobility spectrometry
iTRAQ	isobaric tags for relative and absolute quantitation
K8,19,5,6,14	keratin 8/19/5/6/14
KDM5B	lysine demethylase 5B
KEAP1	kelch like ECH associated protein 1
Ki-67	marker of proliferation Ki-67
KLF4	Krüppel-like factor 4
KLF5	Krüppel-like factor 5
KRAS	KRAS proto-oncogene, GTPase
KS	keratan sulfate
LC	liquid chromatography
LDHA	lactate dehydrogenase A
LFQ	label-free quantification
LKB1	serine/threonine kinase 11, STK11
LMP2A	latent membrane protein 2A
LOESS	locally estimated scatterplot smoothing
LT	large T antigen
LTQ	linear ion trap
macsGESTALT	multiplexed, activatable, clonal and subclonal genome editing of synthetic target arrays for lineage tracing
MAPK	mitogen-activated protein kinase
MCT1	monocarboxylate transporter 1
MEK	mitogen-activated protein kinase kinase (MAP2K1)
MES	mesenchymal
MET	mesenchymal-epithelial transition
MGST1	microsomal glutathione S-transferase 1
miR-200	microRNA 200
miR-27a	microRNA 27a
MMS	mesenchymal metabolic signature

MMTV	mouse mammary tumor virus
MOMP	mitochondrial outer membrane permeabilization
MRS	magnetic resonance spectroscopy
MS	mass spectrometry
MUC	sialomucin
MYC	MYC proto-oncogene, BHLH transcription factor
NAA	<i>N</i> -acetylaspartate
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NAMECs	naturally arising mesenchymal cells
nanoPOTS	nanodroplet processing in one pot for trace samples
NAT8L	aspartate <i>N</i> -acetyltransferase
NEAAs	non-essential amino acids
NF- $\kappa$ B	nuclear factor-kappa B
Non-Mes	non-mesenchymal
NRF2	nuclear factor erythroid 2-related factor 2
NSCLC	non-small cell lung cancer
O <sup>2-</sup>	superoxide
OAA	oxaloacetate
OAT	ornithine aminotransferase
O-GlcNAcylation	O-linked- <i>N</i> -acetylglucosaminylation
O-GlcNAcylation	O-linked- <i>N</i> -acetylglucosaminylation
OGT	O-GlcNAc transferase
OXPHOS	oxidative phosphorylation
p21	cyclin-dependent kinase inhibitor 1
p27	cyclin-dependent kinase inhibitor 1B
p38 MAPK	p38 mitogen-activated protein kinases
PAK1	serine/threonine-protein kinase
PAM50	prediction Analysis of Microarray 50 gene panel
PANK2	isoform 2 of pantothenate kinase 2
PAR1	thrombin receptor
PBS	phosphate-buffered saline
PC	phosphocholine
PCA	principal component analysis
PCR	polymerase chain reaction
PD	proteome discoverer
PDGFR	platelet-derived growth factor receptor
PDHA1	pyruvate dehydrogenase E1 component subunit alpha
PDHks	pyruvate dehydrogenase kinases
PDK-1	pyruvate dehydrogenase kinase 1
PD-L1	programmed death-ligand 1
PFA	paraformaldehyde
PFK	phosphofructokinase
PFKFB3	fructose-2,6-biphosphatase 3
PFKP	ATP-dependent 6-phosphofructokinase
PHGDH	phosphoglycerate dehydrogenase
PI3K	phosphoinositide 3-kinase
PKACA	protein kinase A catalytic subunit $\alpha$
PKC $\alpha$	protein kinase C alpha
PKM2	pyruvate kinase M2
PLA2	phospholipase A2

PLA2G15	phospholipase A2 group XV
PPAR $\gamma$	peroxisome proliferator-activated receptor gamma
PPP	pentose phosphate pathway
PR	progesterone receptor
PRKD1	protein kinase D1
PSM	peptide-spectra-matches
PtdCho	phosphatidylcholine
PtdE	phosphatidylethanolamine
PTMs	post-translational modifications
PVDF	polyvinylidene fluoride
QC-RLSC	quality-control sample-based robust LOESS signal correction
rBM	reconstituted basement membrane
RCD	regulated cell death
ROS	reactive oxygen species
RP	reverse-phase
RPPA	reverse phase protein array
RSD	relative standard deviation
RT	room temperature
RTKs	receptor tyrosine kinases
RT-qPCR	quantitative reverse transcription PCR
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethione
SCoPE-MS	single-cell proteomics by mass spectrometry
scRNA-seq	single-cell RNA-sequencing
SD	standard deviation
SDH	succinate dehydrogenase
SDS	sodium dodecyl sulfate
SILAC	stable isotope labeling by amino acids in cell culture
SIP-1	Smad interacting protein-1
siRNA	small interfering RNA
SIRT1/6	sirtuin 1/6
SIX1	homeobox protein SIX1 (sineoculis homeobox homolog 1)
SLP-2	stomatin-like protein 2
Slug	zinc finger protein SNAI2
SMAD2	mothers against decapentaplegic homolog 2
Snail	zinc finger protein SNAI1
SOD2	mitochondrial superoxide dismutase 2
SORD	sorbitol dehydrogenase
SP1	specificity protein 1
SQOR	sulfide quinone oxidoreductase
Src	proto-oncogene tyrosine-protein kinase Src
SRM	selected reaction monitoring
STAT3	signal transducer and activator of transcription 3
SV-40	simian virus 40
sXBP1	spliced X-Box binding protein 1
TAG	triacylglycerol
TCA	tricarboxylic acid
TCGA	the Cancer Genome Atlas
TDLU	terminal duct lobular units
TEMTIA	the EMT international association
TFA	trifluoroacetic acid

TGF- $\beta$	transforming growth factor- $\beta$
Thy-1	Thy-1 cell surface antigen
TME	tumor microenvironment
TMT	tandem mass tags
TNBC	triple-negative breast cancer
TSTA3	GDP-L-fucose synthetase
TWIST1	twist-related protein 1
UDP-Glc	UDP-glucose
UDP-GlcA	UDP-glucuronic acid
UDP-GlcNAc	UDP-N-acetylglucosamine
UGDH	UDP-glucose dehydrogenase
uPAR	urokinase-type plasminogen activator receptor
UPLC	ultra performance liquid chromatography
VEGFR	vascular endothelial growth factor receptor
YBX1/YB-1	Y-box-binding protein 1
ZEB1/2	zinc finger E-box-binding homeobox 1/2
$\alpha$ -SMA	alpha-smooth muscle actin
$\Delta$ Np63	tumor protein p63, TP63



## List of figures

<b>Figure 1.</b> The EMT and EMP process. ....	10
<b>Figure 2.</b> An overview of the metabolic pathways involved in this project. ....	12
<b>Figure 3.</b> General workflow of LFQ and SILAC proteomics. ....	33
<b>Figure 4.</b> General workflow of a typical metabolomics study. ....	36
<b>Figure 5.</b> An overview of the study workflows in this project. ....	39
<b>Figure 6.</b> Breast spontaneous EMT cell models. ....	62
<b>Figure 7.</b> Comparison of the proteomes among the three breast EMT cell models and during the EMT process. ....	63
<b>Figure 8.</b> Proteomic changes after EMT in the three EMT models. ....	66
<b>Figure 9.</b> Functional changes after EMT in the three EMT models. ....	67
<b>Figure 10.</b> Generation of D492M and D492HER2 from D492 in the EMP cell model. ....	68
<b>Figure 11.</b> The difference to the proteomes among the EMP cell lines. ....	68
<b>Figure 12.</b> Dysregulation of the general and metabolic EMT markers in the EMP cell model and cell line classification of D492, D492M, and D492HER2. ....	71
<b>Figure 13.</b> Proteomic changes among different cell lines in the EMP cell model were quantified by LFQ and SILAC. ....	73
<b>Figure 14.</b> Functional changes among different cell lines in the EMP cell model. ....	75
<b>Figure 15.</b> Consistently altered proteins post-EMT in all three EMT cell models. ....	76
<b>Figure 16.</b> Consistently altered metabolic proteins post-EMT in all three EMT cell models with UGDH altered the most. ....	83
<b>Figure 17.</b> Altered metabolic proteins post two types of EMT processes in the EMP cell model with GFPT2 altered the most in D492HER2 compared with the other two cell lines. ....	85
<b>Figure 18.</b> Functional analysis of the UGDH knockdown in the mesenchymal cells. ....	89

<b>Figure 19.</b> Metabolomics analysis of the UGDH knockdown in the mesenchymal cells. ....	92
<b>Figure 20.</b> The altered GPC and NAA expression with the siUGDH treatment in D492HER2 and MDA-MB-231 and the possible roles of PPAR signaling and PLA2G15 in regulating GPC. ....	95
<b>Figure 21.</b> The upstream signaling regulation of UGDH. ....	98
<b>Figure 22.</b> Functional analysis of the GFPT2 knockdown. ....	99
<b>Figure 23.</b> The GFPT2 expression in different breast cell lines and cancer patients. ....	100
<b>Figure 24.</b> Western blot of protein O-GlycNAcylation with GFPT2 knockdown. ....	102
<b>Figure 25.</b> Metabolomics analysis of the EMP cell model and the GFPT2 knockdown. ....	103
<b>Figure 26.</b> GFPT2 is a marker for oxidative stress. ....	105
<b>Figure 27.</b> The upstream signaling regulation of GFPT2. ....	107
<b>Figure 28.</b> RT-qPCR analysis of GFPT2 with siPKC $\alpha$ treatment. ....	109
<b>Figure 29.</b> Western blot analysis of GFPT2 with siPKC $\alpha$ treatment. ....	109
<b>Figure 30.</b> Reconstruction and upgrade of the D492 GEMs using the proteomic dataset. ....	111
<b>Figure 31.</b> Differently phosphorylated metabolic enzymes between D492M and D492. ....	114
<b>Figure 32.</b> Top differences between D492M and D492. ....	115
<b>Figure 33.</b> Differently phosphorylated metabolic enzymes between D492HER2 and D492. ....	116
<b>Figure 34.</b> Top differences between D492HER2 and D492. ....	117
<b>Figure 35.</b> Differently phosphorylated metabolic enzymes between D492HER2 and D492M. ....	118
<b>Figure 36.</b> Top differences between D492HER2 and D492M. ....	119
<b>Figure 37.</b> Summary of the study on spontaneous EMT. ....	122
<b>Figure 38.</b> Summary of the study on EMP. ....	123

## List of tables

<b>Table 1.</b> Sequences of the siRNAs used in this project.....	52
<b>Table 2.</b> Sequences of the primers used in this project.....	53
<b>Table 3.</b> The composition of the internal standard mix. ....	55
<b>Table 4.</b> The main metabolites identified and quantified in this project were listed. ....	55
<b>Table 5.</b> Consistent proteomic changes in EMT. ....	78
<b>Table 6.</b> Signature proteins in partial mesenchymal D492HER2. ....	79
<b>Table 7.</b> Main metabolic targets dysregulated in the EMP cell model. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.....	86
<b>Table 8.</b> Metabolic pathways affected by UGDH via <i>in silico</i> analysis. ....	93
<b>Table 9.</b> Workflow for analysis of the phosphoproteomic dataset. ....	112
<b>Table 10.</b> Kinases responsible for known phosphosites.....	120



## List of original papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I-III):

- I. Qiong Wang, Sigurdur Trausti Karvelsson, Freyr Johannsson, Arnar Ingi Vilhjalmsson, Lars Hagen, Davi de Miranda Fonseca, Animesh Sharma, Geir Slupphaug, Ottar Rolfsson. (2021) **UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines.** *Molecular Oncology*, doi: 10.1002/1878-0261.13172
- II. Qiong Wang, Sigurdur Trausti Karvelsson, Aristotelis Kotronoulas, Thorarinn Gudjonsson, Skarphedinn Halldorsson, Ottar Rolfsson. (2021) **Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) is upregulated in breast epithelial-mesenchymal transition and responds to oxidative stress.** *Molecular & Cellular Proteomics*, doi: 10.1016/j.mcpro.2021.100185
- III. Sigurdur Trausti Karvelsson, Qiong Wang, Bylgja Hilmarsdottir, Arnar Sigurdsson, Siver Andreas Moestue, Gunhild Mari Mælandsmo, Skarphedinn Halldorsson, Steinn Gudmundsson, Ottar Rolfsson. **Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer.** *NPJ Syst Biol Appl* 7, 36, doi: 10.1038/s41540-021-00195-5.

All papers are reprinted by kind permission of the publishers.

Other publications:

- I. **ECM1 secreted by HER2-overexpressing breast cancer cells promotes formation of a vascular niche accelerating cancer cell migration and invasion.** – [*Laboratory Investigation*, 2020], DOI: 10.1038/s41374-020-0415-6
- II. **EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition.** – [*Molecular Cancer Research*, 2021], DOI: 10.1158/1541-7786.MCR-20-0962

## **Declaration of contribution**

### **Paper #1. UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines**

QW conceived the study, carried out the experiments, performed the data analysis, designed the figures, and wrote the manuscript. STK performed the metabolomic untargeted analysis and the flux analysis in the genome-scale metabolic network reconstructions (GEMs); FJ conducted the metabolomic experiments; AIV carried out experiments; LH, DMF, and AS performed the proteomic analysis under the supervision of GS. OR conceived, supervised, and funded the study, as well as analyzed the data and wrote the manuscript. All authors provided critical feedbacks on the manuscript and data analysis.

### **Paper #2. Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) is upregulated in breast epithelial-mesenchymal transition and responds to oxidative stress**

Conceptualization, QW and OR; Methodology, QW, STK, and AK; Validation, QW; Formal analysis, QW and STK; Investigation, QW; Writing—Original Draft, QW and OR; Writing—Review & Editing, QW, STK, AK, TG, SH, and OR; Visualization, QW; Resources, TG and OR; Supervision, SH and OR; Funding acquisition, OR; Project administration, OR.

We are grateful for the contributions from Douglas Lamont and Amy Tavendale at the 'FingerPrints' Proteomics Facility, College of Life Sciences, MSI/WTB/JBC Complex, University of Dundee for their help with the proteomic and phosphoproteomic data collection and analysis.

### **Paper #3. Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer**

Conceptualization, STK, SG, and OR; Methodology, STK, AR, QW, SAM, SH, SG, and OR; Software, STK and SG; Formal analysis and Investigation, STK, AS, QW, and SAM; Data curation, STK, QW, AS, and SH; Visualization,

STK; Writing—Original Draft, STK and OR; Writing—Review and Editing, all authors; Funding acquisition, GMM, SAM, SG, and OR, Supervision, GMM, SAM, SG, and OR.

**Other publications #1. ECM1 secreted by HER2-overexpressing breast cancer cells promotes formation of a vascular niche accelerating cancer cell migration and invasion**

Steinhaeuser, S. S., Morera, E., Budkova, Z., Schepsky, A., Wang, Q., Rolfsson, O., Riedel, A., Krueger, A., Hilmarsdottir, B., Maelandsmo, G. M., Valdimarsdottir, B., Sigurdardottir, A. K., Agnarsson, B. A., Jonasson, J. G., Ingthorsson, S., Traustadottir, G. A., Oskarsson, T., & Gudjonsson, T. (2020). *Laboratory Investigation*, 1-17. doi:10.1038/s41374-020-0415-6

I facilitated the proteomic data analysis and gave comments on the manuscript.

**Other publications #2. EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition**

Karvelsson, S. T., Sigurdsson, A., Seip, K., Grinde, M. T., Wang, Q., Johannsson, F., Maelandsmo, G. M., Moestue, S. A., Rolfsson, O., & Halldorsson, S. (2021). *Mol Cancer Res*, 19(9), 1546-1558. doi:10.1158/1541-7786.Mcr-20-0962

STK, OR, and SH conceived and designed this study. STK, SH, AS, KS, and MTG performed and analyzed the results of all experiments. QW oversaw running the proteomic method. FJ and OR oversaw running the UPLCMS method. MTG and SAM oversaw running the NMR method. All authors contributed to interpreting the data and edited the manuscript. OR and SH supervised the work carried out in this study. All authors read and approved the final manuscript.





# 1 Introduction

## 1.1 Breast cancer

The burden of cancer is growing fast worldwide. In 2018, the total cost of cancer was 199 billion euros in Europe (Hofmarcher et al., 2020). Around 19.3 million new cases of all cancers and 9.9 million cancer-caused deaths occurred worldwide in 2020. The cancer incidence is expected to rise by 47 %, reaching 28.4 million in the next 20 years, causing a health care burden, especially in transitioning countries. As of 2020, the top estimated age-standardized incidence rate of all cancer types worldwide is breast cancer in females and lung cancer in males. Female breast cancer is the leading cause of cancer deaths in women. In Iceland, female breast cancer had the highest cancer incidence rate in 2020 (Sung et al., 2021).

### 1.1.1 Breast cancer subtypes

Breast cancer is a heterogeneous disease and has been scrupulously stratified based on hormonal phenotypes, including estrogen receptor (ESR) positive, progesterone receptor (PR) positive, receptor tyrosine-protein kinase ERBB2 (HER2) positive, and triple-negative breast cancer (TNBC). On top of the immunohistochemical stratification, breast cancer can be further classified based on genomic and transcriptomic evidence—for example, Ki-67 and breast cancer gene (BRCA1/2) (Loibl et al., 2021). The classic molecular characterization of breast cancer based on the 50-gene PAM50 model has classified it into five intrinsic subtypes: luminal A, luminal B, HER2-enriched, basal-like, and normal-like (Perou et al., 2000; The Cancer Genome Atlas Network, 2012).

Basal-like constitutes most TNBC, which is heterogeneous, considered more aggressive, and has a poorer prognosis than other subtypes (Bianchini et al., 2016). Due to the intrinsic complexity, heterogeneity, and lack of therapeutic targets, TNBC has been a daunting barrier in clinical practice, leaving chemotherapy as the only valid therapeutic option (Denkert et al., 2017). There has been a significant amount of work done to understand the true nature of TNBC and identify effective targets for diagnostics, therapeutics, and prognostics (Denkert et al., 2017; Foulkes et al., 2010). The understanding of TNBC has drastically improved, which leads to more accurate TNBC subtyping for clinical treatment, and one of the breakthroughs

has been the immune-checkpoint inhibitors. In addition, both exploring the crosstalk between tumor cells and tumor microenvironment (TME) and combining immunotherapy and targeted therapy have shown promise in conquering TNBC (Bianchini et al., 2022; Denkert et al., 2017). Nonetheless, more efforts are required and have been made to tackle this problem.

### **1.1.2 Claudin-low breast cancer**

Recently, as an additional molecular subtype of breast cancer, claudin-low has been investigated more thoroughly due to its aggressiveness (Fougner et al., 2020; Prat et al., 2010; Radler et al., 2021). MDA-MB-231, MDA-MB-157, BT549, and HS578T, among others, are widely studied breast cancer mesenchymal cell lines with claudin-low characteristics (Lawrence et al., 2015; Prat et al., 2010) and have been utilized in many studies on claudin-low breast cancer (Patsialou et al., 2015).

Herschkowitz et al. classified and identified claudin-low breast cancer in 2007, seven years after the five intrinsic classifications had been established (Herschkowitz et al., 2007). In that original publication, the claudin-low breast cancer subtype was characterized by low expression of claudins (claudin 3, claudin 4, and claudin 7), occludin, and *E*-cadherin. The sub-classification of the claudin-low subtype has hitherto been open for debate. The biological features represented by claudin-low may coexist with the five intrinsic subtypes. A claudin-low breast tumor may be classified as non-claudin-low, moderately claudin-low, extensively claudin-low, or purely claudin-low (Fougner et al., 2020). Different cell types within the same cellular lineage (mammary stem cells, luminal progenitor cells, or mature luminal cells) can transform into different malignancies triggered by oncogenic insults. Hence, claudin-low breast cancer can be classified into three subgroups based on the cell-of-origin. Two subgroups were related to epithelial-mesenchymal transition (EMT), while the third group was associated with normal human mammary stem cells (Pommier et al., 2020). Radler and colleagues have demonstrated that manipulation of the oncogene KRAS could induce claudin-low mammary cancer in luminal epithelial cells (Radler et al., 2021).

### **1.1.3 Breast cancer metastasis**

Metastatic breast cancer accounts for the poor patient prognosis, and the cure for breast cancer metastasis is elusive. The combination of surgery, radiation therapy, and chemotherapy, along with targeted approaches and immunotherapy, is the current standard of care in the clinic (Loibl et al., 2021). The metastatic breast cancer cells are considerably different from

primary tumor cells concerning their genetic landscape, the potential to interact with the immune system and their TME, and cell-type diversity (Bertucci et al., 2019; Hutchinson et al., 2020). Studies show that circulating tumor cells (CTCs) have significantly enriched with cells that have undergone EMT, and the tumor cell population is associated with different transition states of EMT, rendering different potential in cancer metastasis (Pastushenko et al., 2018). EMT may facilitate metastatic dissemination epigenetically without somatic mutations in primary tumor cells (Q.-L. Liu et al., 2021).

## **1.2 EMT**

### **1.2.1 The definition of EMT**

In 2020, The EMT International Association (TEMTIA) published a consensus statement intending to standardize EMT research and reduce discrepancies and misinterpretations due to the plasticity and heterogeneity of EMT (J. Yang et al., 2020).

A multifaceted and often reversible change in cellular phenotypes during which epithelial cells lose their apical-basal polarity, modulate their cytoskeleton and exhibit reduced cell-cell adhesive properties. Cells may individually or collectively acquire mesenchymal features and increase motility and invasive ability. Typically, a switch in intermediate filament usage from cytokeratins to vimentin is observed after a complete EMT. Cortical actin filament in epithelial cells also undergoes marked rearrangement during EMT. While the characteristics of fully epithelial cells are relatively clearly defined, our current knowledge does not allow us to define the mesenchymal state with specific cellular characteristic or molecular markers that are universal end products of all EMT programs. (J. Yang et al., 2020)

There are two main take-home messages from this guideline. First, EMT cannot be simply defined based on one or several molecular markers or EMT-associated transcription factors (EMT-TFs). Second, EMT should be assessed on the basis of both changes in the cellular characteristics and a set of molecular markers (J. Yang et al., 2020).

## 1.2.2 The history of EMT

### ***Origin of the EMT concept***

Although the observation and description of the interconversions that take place between the epithelial and mesenchymal cell states during embryonic development can be traced back to the beginning of the 20<sup>th</sup> century, the origin of the EMT theory dates back to 1968, when Dr. Elizabeth Hay gave a speech at the 18<sup>th</sup> Hahemann symposium in Baltimore about her embryogenic research (Hay, 1968). Later, Dr. Hay and colleagues manipulated the adult and embryonic epithelia *in vitro* and successfully transformed epithelium into mesenchymal phenotypes, a phenomenon that defined EMT as a distinct cellular process (Greenburg and Hay, 1982). In the 20 years that followed, several key features of EMT were defined, including the loss of *E*-cadherin; the gain of *N*-cadherin; and the importance of specific transcription and growth factors in EMT, such as Snail, Slug, E2A, ZEB1/2, fibroblast growth factor (FGF), insulin-like growth factor (IGF), transforming growth factor- $\beta$  (TGF- $\beta$ ), epidermal growth factor (EGF), and ERBB (Thiery, 2002). In 1991, a pioneer study highlighted *E*-cadherin as an invasion suppressor and connected the mesenchymal phenotype to invasiveness, which paved the way for the hypothesis that EMT can be exploited by metastatic cancer cells (Frixen et al., 1991).

Mesenchymal-epithelial transition (MET) is a reverse process of EMT, during which mesenchymal cells regain their apical-basal polarity and rearrange their cytoskeleton and cell-cell adhesion to form an organized epithelium. Compared with EMT, MET has been studied less in cancer research. It is well defined in kidney development and the formation of the heart and somites (Pei et al., 2019; Thiery, 2002; J. Yang et al., 2020). The existence of MET is primarily evident by the restoration of *E*-cadherin, and the MET process has been confirmed in breast cancer, ovarian cancer, and pancreatic cancer via probing specific epithelial or mesenchymal markers (Bakir et al., 2020). At the beginning of the 21<sup>st</sup> century, a novel and intriguing theory of cancer was established that incorporated the EMT phenomenon into cancer metastasis. This theory is illustrated as: normal epithelium  $\rightarrow$  dysplasia/adenoma  $\rightarrow$  carcinoma *in situ*  $\rightarrow$  invasion carcinoma via EMT  $\rightarrow$  intravasation via EMT  $\rightarrow$  extravasation via EMT  $\rightarrow$  formation of distant carcinoma via MET (Thiery, 2002). In 2003, the term “epithelial-mesenchymal transition” was officially coined at the first TEMENTIA meeting. Since then, research on EMT has grown drastically, especially over the last 10 years (J. Yang et al., 2020).

### ***The classical molecular components in EMT***

The downregulation of *E*-cadherin in EMT was first observed in the embryonic development of chicks (Edelman et al., 1983), followed by studies on the loss of *E*-cadherin in cancer (Frixen et al., 1991). On the contrary, *N*-cadherin was observed to be upregulated in EMT during embryogenesis, though at different sites of the embryo compared with the downregulation of *E*-cadherin (Duband et al., 1987). Even though the switch from *E*-cadherin to *N*-cadherin has become a hallmark of EMT (Loh et al., 2019), loss of *E*-cadherin has emerged as an essential and central characteristic in EMT (Lachat et al., 2021). As indicated in the EMT definition, high expression of the mesenchymal intermediate filament vimentin is another pivotal change in EMT, which has long been recognized in embryogenic EMT (Greenburg and Hay, 1988) and is associated with the invasive mesenchymal traits in breast cancer (Bae et al., 1993).

In the 1990s, a large-scale study set out to identify molecular regulators of EMT in embryonic development. Several well-established EMT-TFs were identified around the turn of the millennium—for example, Snail, Slug, TWIST1, ZEB1, and ZEB2 (Thiery, 2002). Later, these EMT-TFs were also found to be involved in cancer malignancy (Batlle et al., 2000; Cano et al., 2000; Comijn et al., 2001; J. Yang et al., 2004). In 1994, EMT was successfully induced by TGF- $\beta$  *in vitro* (Miettinen et al., 1994). Ectopic expression of the EMT-TFs Snail or TWIST in the epithelial cells was found to trigger EMT (Mani et al., 2008).

#### **1.2.3 EMT in cancer**

EMT is mainly studied in tissue development, wound healing and organ fibrosis, and cancer progression. Despite the significance of EMT in embryogenesis and fibrogenesis, most EMT studies conducted so far have focused on cancer aggressiveness, specifically cancer stem cell (CSC) formation, metastasis, and chemoresistance. The plasticity of EMT provides comprehensible explanations for tumor extravasation and distant organ colonization during cancer metastasis (Lachat et al., 2021).

Researchers have found evidence to support that both complete and partial EMT exist *in vivo* (Beerling et al., 2016; Pastushenko et al., 2018). EMT is highly appreciated in epithelial-tissue-derived tumors (carcinomas), and studies have indicated that EMT also occurs in tumors of the central nervous system (CNS) and perhaps other non-carcinomas (Wirsik et al., 2021). The biopsies of cancer patients show various cell phenotypes ranging

from epithelial to mesenchymal-like, and the cellular phenotypic plasticity causes cancer cells to be resistant to drug treatment (Navas et al., 2020). Increasing evidence suggests that EMT participates in both the initiating and the final stages of cancer metastasis. Cells at the leading edges of the invasive fronts could capture the EMT properties and lead the invasion process (Wellner et al., 2009). EMT may promote cell stemness. The association between EMT and cancer invasion and metastasis has been widely studied. However, the exact mechanisms behind EMT in stemness and metastatic dissemination are obscure (Lambert and Weinberg, 2021).

EMT-TFs are a crucial part of the EMT program and participate in cancer initiation, tumor growth, invasion, metastasis, and colonization. In addition to the roles of these transcription factors on the activation of the classical EMT phenotypes, they are pleiotropic and can link EMT to CSCs. The uncanonical functions of EMT-TFs can facilitate cancer malignancy, therapy resistance, TME crosstalk, and tumor immune response. Lastly, the EMT-TFs grant EMT plasticity (Brabletz et al., 2018; Stemmler et al., 2019).

EMT is very diverse, tissue-specific, cancer type-specific, and intertwined with many regulatory factors and signaling pathways. Some studies have reported that metastasis *in vivo* occurs without EMT (Bakir et al., 2020; Fischer et al., 2015; Williams et al., 2019). EMT in cancer is debatable, and the source of this controversy originates from three intrinsic and fundamental characteristics of EMT and EMT research: complexity, plasticity, and suboptimal experimental models. EMT may manifest itself differently in different organs or distinct cancer types. The conclusion that EMT is absent in metastatic sites based on the absence of overt mesenchymal characteristics is insufficient because certain levels of reversion to the epithelial state are necessary. In addition, EMT impinges both on cell metastasis and on cell survival—for example, chemoresistance. Scientists need to be open-minded, considering EMT beyond its effects on cell morphology, invasion, and motility (Brabletz et al., 2018).

One of the future directions is to focus on EMT-TFs and their interactions with the TME. Furthermore, more reliable experimental models, for example, genetically engineered mouse models (GEMMs) coupled with lineage tracing with fluorescent labeling, should be applied to tackle the controversies and inconsistencies in this field (Bakir et al., 2020). Researchers need to collaborate closely and perform cross-validations of their findings. Moreover, the challenges of monitoring all EMT states and following all cancer cells from cancer initiation to distant colonization with a focus on the metastatic

niche should be overcome. The final goal is to identify clinically meaningful targets (Brabletz et al., 2018).

### ***The intertwined relationship between EMT and CSCs***

Cancer has long been associated with stem cells, consolidated by the observation that a single cell could form tumors *in vivo*. The successful identifications of CSCs in acute myeloid leukemia (AML) and breast cancer are two breakthroughs that created great excitement in the cancer research community (Clevers, 2011). Since then, the concept of CSCs has been accepted and updated. One of the novel improvements of this theory is the introduction of the CSC niche. Instead of being solely regulated by the intrinsic properties of CSCs, CSCs interact via niche signaling (Batlle and Clevers, 2017). One of the appealing aspects of the CSC hypothesis is that it explains the inevitable resistance of cancer cells to radiation and chemotherapy and the long-term relapse and metastasis in patients with cancer. Under the guidance of the CSC theory, several therapeutic approaches have been developed in the clinic by targeting major stemness pathways, the main stem cell properties, and epigenetic regulations, among other factors (Batlle and Clevers, 2017).

The association between EMT and CSCs was initiated through research focusing on how exactly EMT facilitates tumor metastasis. It began with the identification of CD44<sup>high</sup>/CD24<sup>low</sup>-expressing cells as CSCs (Al-Hajj et al., 2003). These cells are more invasive *in vivo* and possess stem cell properties. Mani and colleagues then successfully induced EMT by TGF- $\beta$  treatment and ectopic expression of TWIST or Snail, and these treatments simultaneously led to the evolution of a population of CD44<sup>high</sup>/CD24<sup>low</sup> cells (Mani et al., 2008). These efforts linked EMT and CSCs and sparked interests to investigate further whether EMT could facilitate CSC formation in cancer progression (Wilson et al., 2020). CSCs are intertwined with EMT through shared genetic regulators and tumor-initiating capacity, and EMT can confer epithelial cells with stem cell properties (Mani et al., 2008). In the context of cancer, EMT was initially deemed a vital process for cancer migration and invasion. Connecting EMT with CSCs has established the role of EMT in cancer initiation and has intensified the importance of EMT in cancer development. The mesenchymal cells after EMT and CSCs share lots of similarities (Dongre and Weinberg, 2019; Wilson et al., 2020). The exact relationship between CSCs and EMT, however, is still under debate (Batlle and Clevers, 2017; Lytle et al., 2018).

Epidermal growth factor receptor (EGFR) is liable for the CSC and metastatic phenotype of breast cancer, and platelet-derived growth factor receptors (PDGFRs) promote breast cancer malignancy by directly regulating cellular functions and by indirectly remodeling malignant stroma (Butti et al., 2018). The exact relationship between EGFR and CSCs is elusive. Tam and colleagues reported a shift from EGFR to PDGFR signaling in breast CSCs, where protein kinase C alpha (PKC $\alpha$ ) participates and serves as a therapeutic target (Tam et al., 2013). PKC $\alpha$  is universally expressed and situated at the central node of several signal transduction pathways, which endows the diverse and complex cellular responses of PKC $\alpha$  under a broad spectrum of stimuli (Singh et al., 2017).

***Epithelial-mesenchymal plasticity (EMP) represents the norm rather than the exception***

The concept of “partial EMT” was proposed roughly 20 years ago (Thiery, 2002), and EMP is one of the sources for discrepancies in the EMT field. In the time since, there has been compelling evidence that EMT does not work as a binary switch but possesses a spectrum of epithelial/mesenchymal (E/M) intermediate states within which certain E/M states confer the tumor with cells malignancy (Liao and Yang, 2020; Sinha et al., 2020) (**Figure 1**). These varieties of intermediate states are orchestrated by complex regulatory networks that endow them with different functional characteristics (Dongre and Weinberg, 2019; Nieto et al., 2016; Pastushenko and Blanpain, 2019). It is believed that EMT can induce CSC stemness. While these CSCs induced by EMT are not completely mesenchymal, they reside and are stabilized at specific intermediate E/M states (Dongre and Weinberg, 2019). Partial EMT can increase cell motility in both single and collective cell migration.

Recent studies show that different EMT transition states do exist *in vivo* (Pastushenko et al., 2018; Simeonov et al., 2021). There were different subpopulations of tumor cells *in vivo* that possessed different invasiveness, plasticity, and functionality, representing different EMT states. Furthermore, the metastatic potential *in vivo* was affected by different hybrid EMTs where tumor protein p63 ( $\Delta$ Np63), TGF- $\beta$ , and mothers against decapentaplegic homolog 2 (SMAD2) were important promoters for epithelial or mesenchymal phenotypes, and MET was not the only mechanism for metastasis (Pastushenko et al., 2018). More recently, a novel inducible *in vivo* lineage tracer coupled with scRNA-seq named macsGESTALT was introduced to study EMT plasticity, which confirmed the notion that EMT is a continuum of epithelial-mesenchymal states. Cells post extreme EMT were less metastatic



than hybrid EMT cells, and those at the late-stage hybrid states showed the most potential to metastasize (Simeonov et al., 2021).

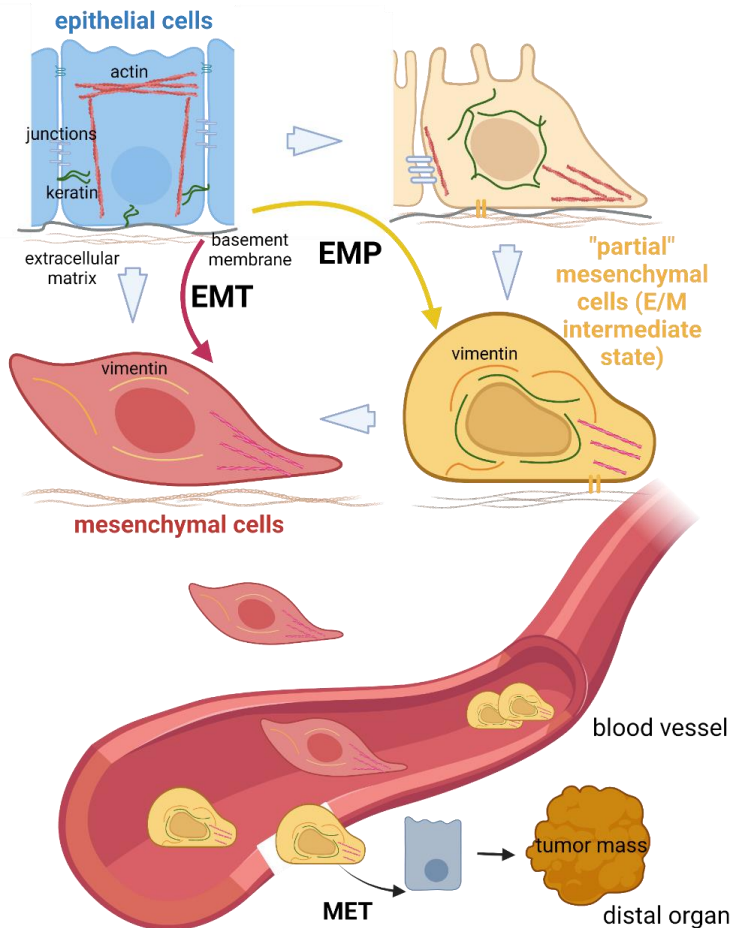
Apart from relying on specific epithelial or mesenchymal markers (for example, CD104 and CD44) to differentiate different EMT phenotypes (Kröger et al., 2019), *in silico* modeling of hybrid EMT provides an alternative for investigating the complexity and plasticity of the EMT spectrum. More evidence is needed to support whether discrete epithelial and mesenchymal states are arrayed along the EMT spectrum or a continuum of the EMT intermediate states resides with no distinct boundaries. The EMT spectrum may not be linear—it might have various tracks. Sophisticated mathematical models hold the potential to tackle these problems and are able to forge new insights in this field, and a list of studies using this methodology has demonstrated the importance of EMT plasticity in cancer metastasis, resistance to therapy, stemness, and immune responses (Jia et al., 2019; Jolly et al., 2019; Lu et al., 2013; Tripathi et al., 2020).

### ***Collective cell migration: A promising theory for cancer invasion***

Single-cell migration is not the only mechanism for cancer cell invasion. Collective cancer cell invasion is also important for disseminating tumor cells to distal sites; that is, cells invade the peritumoral stroma without losing cell-cell contacts. The grouped CTCs in patients' peripheral blood demonstrate that collective tumor migration exists *in vivo* (Aceto et al., 2014; J. M. Hou et al., 2011). These cohesive multicellular clusters in breast cancer could be abrogated by inhibiting plakoglobin, keratin 14, or p63 (Aceto et al., 2014; Cheung et al., 2013). CTCs isolated from patients with breast cancer possess many mesenchymal characteristics, and these mesenchymal CTCs are associated with the patient's disease progression and responses to therapy (M. Yu et al., 2013). The mesenchymal states of CTCs in patients may reflect cancer progressiveness and hold prognostic value (Williams et al., 2019).

The core of the EMT concept is the function of the EMT cellular program in transforming adhesive cells into non-adhesive migrating cells, which has been primarily investigated through single-cell migration and invasion assays. However, EMT occurs in both a single cell format and a collective of cells (M. Yu et al., 2013). The "leader cells" at the tip of the invading cell groups could undergo transient or partial EMT induced by microenvironmental cues to guide collective cancer cell invasion (Friedl et al., 2012). By adopting a microfluidic approach, Wong and colleagues monitored single-cell dynamics during cell migration. They reported that cells formed a collective advancing

front during migration via EMT, and there are single cells scattered away from these cell clusters (Wong et al., 2014).



**Figure 1.** The EMT and EMP process.

Epithelial-mesenchymal transition (EMT) is a natural cellular process that can be hijacked by cancer cells during cancer metastasis. In the EMT process, epithelial cells go through cytoskeleton remodeling, lose the apical-basal cell polarity, weaken cell-cell and cell-matrix adhesions, and acquire cell motility and the ability to invade the basement membrane to form mesenchymal cells. In most cases, epithelial cells undergo a certain degree of mesenchymal transition to form "partial" or intermediate mesenchymal cells, a phenomenon referred to as epithelial-mesenchymal plasticity (EMP). These mesenchymal cells are generally more aggressive and can invade and metastasize via extravasation through blood or lymphatic vessels to distal organs (in a single-cell form or collectively). Mesenchymal cells transform back to the epithelial phenotype at distal organs via mesenchymal-epithelial transition (MET), followed by settlement and proliferation.

### 1.2.4 The therapeutic implications of EMT

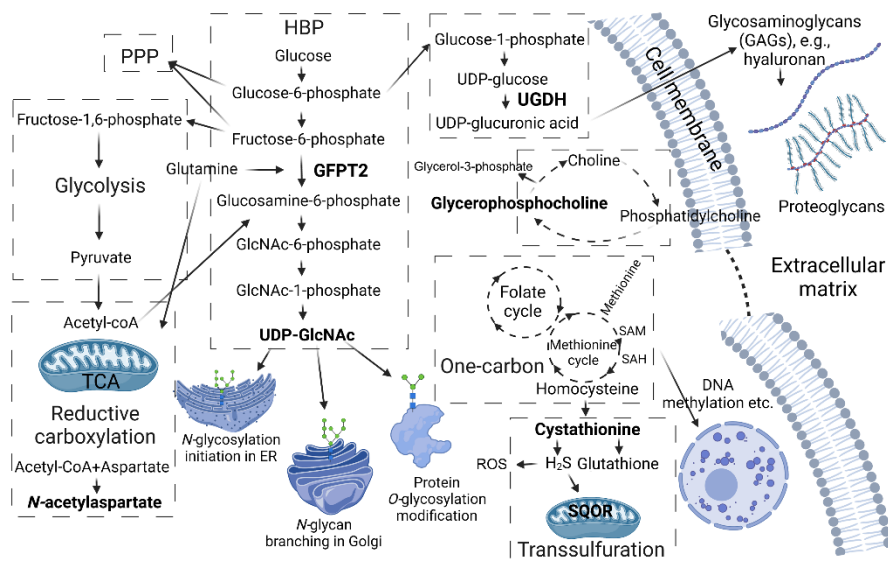
As of 2022, there have been 54 clinical trials registered on [clinicaltrials.gov](https://clinicaltrials.gov) that directly or indirectly target EMT as therapeutic approaches. Eleven trials have breast cancer as one of the targeted conditions. The intrinsic properties of partial EMT have created a formidable challenge for EMT-targeting therapy due to the difficulties in identifying a druggable intermediate E/M state. Targeting the EMT-related modulators, reversing EMT, and inhibiting mesenchymal-like cells represent the three main strategies (Voon et al., 2017).

EMT-related modulators nuclear factor-kappa B (NF- $\kappa$ B), signal transducer and activator of transcription 3 (STAT3), hypoxia-inducible factor-1 $\alpha$  (HIF-1 $\alpha$ ), and TGF- $\beta$  are four examples out of many potential targets of EMT regulatory components (Voon et al., 2017; X. G. Yang et al., 2019). Targeting specific EMT-related cancer-inducing regulators will eventually cause drug insensitivity. EMT is potentially a pathological manifestation of somatic cell reprogramming that renders cellular plasticity, which is accountable for chemoresistance (Voon et al., 2017). CSCs with mesenchymal characteristics have been deemed the culprit for cancer chemoresistance. Inhibiting mesenchymal-like CSCs could potentially prevent the resistance to chemotherapy (Voon et al., 2017). Reversing the invasive mesenchymal cells back to the epithelial-like phenotype is another rationale of EMT therapy. This includes inhibitors on ALK5/TGF $\beta$ R1, mitogen-activated protein kinase (MAPK), proto-oncogene tyrosine-protein kinase Src, focal adhesion kinase (FAK), and phosphoinositide 3-kinase (PI3K) for their MET-inducing or anti-EMT activities. However, precautions should be taken regarding EMT-reversing therapy because MET is involved in distant colonization during metastasis (Voon et al., 2017).

The interaction between EMT and the immune responses has been under intense scrutiny (Dongre and Weinberg, 2019). Of note, the EMT-related signaling pathway miR-200/ZEB1 has been shown to increase PD-L1 expression in tumors, and EMT-TFs TWIST can recruit macrophages, corroborating the immunosuppressive effects of EMT (L. Chen et al., 2014; Low-Marchelli et al., 2013). Targeting EMT might exert synergistic effects on immunotherapy. Altered metabolism is one of the main features of EMT (Sciacovelli and Frezza, 2017). Glycolysis, lipid metabolism, mitochondrial metabolism, pyrimidine metabolism, and more are potential candidates to intervene in the EMT process. Targeting EMT-related metabolic pathways is another alternative for EMT-associated drug therapies (Ramesh et al., 2020).

### 1.3 Cancer metabolism

In the following section, I will first generally review metabolic reprogramming in cancer and breast cancer, cancer metastasis and EMT, and the interaction of metabolism and signaling regulation. Then, I will focus on aspects related to this project, specifically glycan metabolism, including two enzymes studied in-depth: UDP-glucose dehydrogenase (UGDH) and glutamine fructose-6-phosphate transaminase 2 (GFPT2), one-carbon metabolism, the transsulfuration pathway, reactive oxygen species (ROS) regulation, mitochondrial metabolism, and glycerophosphocholine (GPC) and *N*-acetylaspartate (NAA) metabolism (**Figure 2**).



**Figure 2.** An overview of the metabolic pathways involved in this project.

In the hexosamine biosynthetic pathway (HBP), glucose is finally converted into UDP-*N*-acetylglucosamine (UDP-GlcNAc), indispensable for protein glycosylation modification and glycan formation. Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) is a rate-limiting enzyme of the HBP. The HBP is intertwined with the pentose phosphate pathway (PPP), glycolysis, and the formation of precursors for glycosaminoglycans (GAGs) and proteoglycans, essential components in the extracellular matrix (ECM). UDP-glucose dehydrogenase (UGDH) is responsible for converting UDP-glucose to UDP-glucuronic acid, one of the building blocks of hyaluronan in ECM. The product of glycolysis, pyruvate, is converted into acetyl-CoA in the mitochondria, which is a substrate in the HBP, along with glutamine. Glutamine can be metabolized in the mitochondria to generate acetyl-CoA and aspartate via reductive carboxylation, and these two metabolites are then converted into *N*-acetylaspartate (NAA). The other two pathways related to this project are choline metabolism, in which glycerophosphocholine (GPC) is converted to free choline and glycerol-3-phosphate, and the transsulfuration pathway, which is associated with one-carbon metabolism. Homocysteine participates in the methionine cycle in one-carbon

metabolism, within which active methyl-donors are produced. Homocysteine can also be converted into cystathionine to generate hydrogen sulfate ( $H_2S$ ) and glutathione in the transsulfuration pathway, both of which are involved in reactive oxygen species (ROS) regulation. Sulfide quinone oxidoreductase (SQOR), an enzyme that binds to the mitochondrial membrane, uses  $H_2S$  as substrates.

### 1.3.1 Cancer and breast cancer metabolism

Metabolism comprises a series of chemical reactions that occur within living cells to provide energy, to produce building blocks, to regulate biological processes, and to facilitate cell-cell and cell-extracellular matrix (ECM) interactions. Metabolic dysregulation is deeply intertwined with oncogenesis (Counihan et al., 2018). Metabolic reprogramming as an emerging hallmark of cancer can be grouped into seven categories: (1) altered glucose and amino acids uptake; (2) acquired flexibility for nutrient demand; (3) increased utilization of the intermediates from glycolysis and the tricarboxylic acid (TCA) cycle for biosynthesis and the production of nicotinamide adenine dinucleotide phosphate (NADPH); (4) elevated demand for reduced nitrogen due to the increased cell growth; (5) alterations in gene expression regulated by, for example, acetylation and methylation; (6) exchange with the TME; and (7) maintaining redox balance (Pavlova and Thompson, 2016).

Cancer cells produce energy via glycolysis instead of aerobic respiration in the mitochondria, resulting in more lactic acid production than normal cells, a phenomenon known as the “Warburg effect,” observed by Otto Warburg in 1927, which first linked metabolism to oncogenesis (Warburg et al., 1927). Apart from aerobic glycolysis, the influences of mitochondrial metabolism have started to earn appreciations in carcinogenesis (Porporato et al., 2018). The discovery that fructose-2,6-bisphosphate allosterically activates phosphofructokinase (PFK) has revolutionized the common understanding of metabolites in cellular regulations (Van Schaftingen et al., 1981). To date, numerous metabolic enzymes have been implicated in cancer for their canonically enzymatic and uncanonically regulatory functions (Martinez-Reyes and Chandel, 2021). In addition, cancer-associated fibroblasts (CAFs) are the most abundant intratumoral cell type to support malignant tumor cell growth, survival, and metastasis (Kalluri, 2016). The tumor-promoting effects of CAFs are dependent on the production of lactate, counterbalancing the TME acidification, activation of autophagy to supply cancer cells with non-essential amino acids (NEAAs), and regulating glucose uptake (Schworer et al., 2019; W. Zhang et al., 2018). These changes agree with the “reverse Warburg effect” proposed in 2009, stating that carcinoma cells stimulate the Warburg effect or aerobic glycolysis in neighboring stromal CAFs that secrete

many energy-rich metabolites, including lactate and pyruvate to “feed” the tumor cells (Pavlidis et al., 2009).

### ***Metabolic reprogramming in breast cancer***

The metabolic abnormality of breast cancer has been extensively studied and reviewed (D. Zhang et al., 2021). Both primary and metastatic breast cancer cells exhibit metabolic heterogeneity (L. Wang et al., 2020). Different breast cancer subtypes present distinct metabolic phenotypes (D. Zhang et al., 2021).

In TNBC, EGF signaling reprograms glycolysis to accumulate the intermediate fructose 1,6 bisphosphate (F1,6BP), which can bind to and enhance EGFR activity, leading to augmented aerobic glycolysis, increased lactate secretion, and impaired immunosurveillance mediated by cytotoxic T cells (S. O. Lim et al., 2016). The transcription factor MYC is disproportionately overexpressed in TNBC compared with the other types, and fatty acid oxidation intermediates are significantly upregulated for energy metabolism in MYC-overexpressing TNBC (Camarda et al., 2016). ESR-negative breast cancer cells overexpress phosphoglycerate dehydrogenase (PHGDH) in the serine biosynthesis pathway to fuel the anaplerosis of the TCA (Possemato et al., 2011). HER2-positive breast cancer possesses high glutamine metabolic activities (S. Kim et al., 2013).

The energy sensor AMP-activated protein kinase (AMPK) and fatty acid synthase (FASN) have also received considerable attention as therapeutic targets for breast cancer treatment (W. Cao et al., 2019; Menendez and Lupu, 2017). HIF signaling regulates glucose, amino acids, lipids, and ROS metabolism in breast cancer, which can be exploited in combination with immune- and endocrine therapy in the clinic (de Heer et al., 2020). In addition, hexokinase (HK), pyruvate dehydrogenase kinase 1 (PDK-1), and glutaminase, among other enzymes, are promising metabolic targets in breast cancer treatment and are being evaluated in clinical trials (L. Wang et al., 2020). To date, however, the definitive clinical benefits of interfering with metabolism for/alongside breast cancer treatment are lacking (L. Wang et al., 2020).

### **1.3.2 Metabolic reprogramming in metastasis and EMT**

In the cancer metastatic cascade, metabolic rewiring has emerged as an important aspect relying on metabolic plasticity and flexibility. The metabolic phenotypes between primary tumor cells and cells that have undergone

metastasis and those of different metastatic sites are different (Bergers and Fendt, 2021). Pyruvate, lactate, glutamine, fatty acids, and more are common metabolites with functional plasticity in cancer metastasis, all of which are intertwined with various enzymes and signaling regulators and participate in cell invasion, circulation, and colonization in the metastatic cascade (Christen et al., 2016; Rodrigues et al., 2016; Tasdogan et al., 2020). Cancer cells depend on different metabolites to fulfill the same metabolic needs, reflecting their metabolic flexibility. Metastasized cancer cells that seeded at distant sites demand more energy production (Elia et al., 2018). Fatty acids, glucose, proline, lactate, and the ATP-scavenging machinery have been reported to accommodate the excess ATP need in cancer cells colonizing distant sites (Dupuy et al., 2015; Elia et al., 2017; Loo et al., 2015; J. H. Park et al., 2016).

The metabolic rewiring of cancer cells to accommodate the needs for cell proliferation, circulation, seeding, and adaptation shows metabolic plasticity and flexibility, but concurrently, it also exposes certain metabolic rigidity and specific vulnerabilities. For example, blocking proline catabolism and inhibiting lactate uptake dramatically affect cancer metastasis (Tasdogan et al., 2020). This has led to promising metabolic therapeutic strategies to target specific metabolic vulnerabilities depending on different cancer progression phases of patients.

Cancer cells require the pentose phosphate pathway (PPP) to meet the anabolic demands for high proliferation and to resist oxidative stress for drug resistance. The production of NADPH in the PPP is compensated by AMPK-mediated NADPH generation under energetic stress, both of which are pivotal for cancer cells to overcome oxidative stress during metastasis (Patra and Hay, 2014).

### ***Metabolic reprogramming in EMT***

In carcinogenesis, metabolic rewiring and EMT are entwined (Morandi et al., 2017; Sciacovelli and Frezza, 2017). EMT leads to broad metabolic reprogramming covering but not limited to mitochondrial metabolism; hypoxia; and glucose, glutamine, serine, fatty acid, and nucleic acid metabolism (M. Li et al., 2019). Shaul and colleagues identified 44 metabolic genes associated with mesenchymal features, designated the “mesenchymal metabolic signature” (MMS), covering nucleotide, lipid, amino acid, carbon, redox, glycan, cofactor, and other metabolic pathways. They reported that the pyrimidine-degrading enzyme dihydropyrimidine dehydrogenase (DPYD) and its product dihydropyrimidines were responsible for the mesenchymal traits

and malignant transformation (Shaul et al., 2014). TGF- $\beta$ -induced EMT has considerable impacts on both glycolysis and mitochondrial metabolism by deregulating several related enzymes, such as GLUT1, HK2, fructose-2,6-biphosphatase 3 (PFKFB3), pyruvate kinase M2 (PKM2), LDHA, PDK-4, fumarate hydratase (FH), succinate dehydrogenase (SDH), and IDH (Hua et al., 2020). The upregulation of xenobiotic-metabolizing enzymes, including drug transporters, cytochrome P450s, and glutathione-related enzymes, after EMT renders cancer cells chemoresistance (Fischer et al., 2015).

EMT can also be shaped by metabolic changes. Increased glycolytic activity results in higher production of lactate, which regulates the NAD<sup>+</sup>/NADH ratio to affect the NAD<sup>+</sup>-dependent enzyme, sirtuin 1 (SIRT1), that induces EMT (Eades et al., 2011). In addition, the acidic microenvironment has been found to induce EMT (Morandi et al., 2017). Hypoxia activates several EMT-TFs (including Snail1, Slug, and TWIST), with HIF-1 $\alpha$  playing a critical role (J. Jiang et al., 2011). Targeting glutaminase 1 (GLS1) to impair glutamine metabolism can also disrupt the EMT program and hamper tumor growth and metastasis (Morandi et al., 2017). Fatty acids are cellular fuel, building blocks in cell membranes, and signaling molecules. Fatty acid-binding protein (FABP5) can induce EMT to facilitate metastasis in hepatocellular carcinoma cell lines (Ohata et al., 2017). Inhibition of lysine demethylase 5B (KDM5B) can activate AMPK to downregulate FASN and ATP citrate lyase (ACLY) in fatty acid metabolism, reverse EMT, and eventually inhibit breast cancer cell proliferation and migration (Z. G. Zhang et al., 2019). The role of FASN is controversial: FASN can be downregulated by EMT induction, a change that is inconsistent with the active FASN-driven lipogenesis in EMT (Morandi et al., 2017). Moreover, the fatty acid species, sphingolipids, and lipid composition, including cholesterol content, are affected by EMT and concomitantly shape the EMT program (Eiriksson et al., 2018; Morandi et al., 2017). SDH5, one of the subunits in the SDH complex of the electron transport chain (ETC), activates glycogen synthase kinase 3 beta (GSK-3 $\beta$ ) to suppress Wnt signaling, which in return affects EMT (J. Liu et al., 2013). Mitochondrial superoxide dismutase 2 (SOD2) in ROS regulation is upregulated after TGF- $\beta$ -mediated EMT, and partially impaired mitochondrial metabolism is associated with ROS and worse patient survival in different types of cancer and correlated with the EMT gene signature (Eduardo Gaude and Frezza, 2016; Morandi et al., 2017).



### 1.3.3 Crosstalk between oncogenic signaling and metabolism

Cell metabolism is regulated by a complex network of signaling pathways partially mediated via receptor tyrosine kinases (RTKs), such as EGFR, IGF1R, PDGFR, and vascular endothelial growth factor receptor (VEGFR). They play crucial roles in breast cancer progression (Butti et al., 2018). Activation of IGF1R results in the activation of PI3K/AKT and MAPK, which leads to ESR downregulation and resistance to tamoxifen (Schiff et al., 2004; Vella et al., 2020). PI3K/AKT signaling is intimately related to metabolism (Counihan et al., 2018), and PI3K-AKT1 drives the tumor glycolytic phenotype (Hoxhaj and Manning, 2020).

In return, metabolic changes modulate signaling activities and gene expression that drive cell migration and invasion (Elia et al., 2018). Enzymes may exceed their catalytic roles and exert nonconventional signaling regulatory effects, which complicates the role of metabolic enzymes in cellular regulations. GSK-3, which was initially identified in glycogen metabolism, plays a key role in Wnt signaling and insulin regulation (Cohen and Frame, 2001). Insulin-induced activation of PDK-1 and AKT leads to the phosphorylation of GSK-3 with the help of PI3K, eventually causing the suppression of GSK-3 activity and subsequent stimulation of glycogen synthesis (Cohen and Frame, 2001). NF- $\kappa$ B is a well-known transcription factor under the regulation of PI3K/AKT and is essential for EMT induction and maintenance (Huber et al., 2004). The monocarboxylate transporter 1 (MCT1), which exchanges lactate across the plasma membrane, has been shown to activate NF- $\kappa$ B and promote tumor growth and metastasis, independent of its lactate transport function (Z. Zhao et al., 2014). Fatty acid metabolism is also intimately related to the signaling regulation of EMT. Elevated free fatty acid uptake activates the Wnt and TGF- $\beta$  signaling pathways to induce EMT and promote cancer progression (Nath et al., 2015). The unsaturated lipids are significantly increased in ovarian cancer stem cells. NF- $\kappa$ B regulates lipid desaturases and the inhibition of which blocks NF- $\kappa$ B signaling involved in EMT (J. Li et al., 2017).

### 1.3.4 Metabolic reprogramming in glycan metabolism

The flux through the hexosamine biosynthetic pathway (HBP) and the level of UDP-*N*-acetylglucosamine (UDP-GlcNAc) are upregulated in different cancer types. The HBP is closely associated with glycolysis and accounts for 2%-5% of total glucose metabolism (Akella et al., 2019; Marshall et al., 1991). Glucose, glutamine, fatty acids, and amino acids are substrates of the HBP and the main metabolic drivers of tumor growth. Their availability alters the

HBP flux and the production of UDP-GlcNAc, supporting the nutrient-sensing function of the HBP for metabolic homeostasis (Chiaradonna et al., 2018).

The end-product of HBP, UDP-GlcNAc, is indispensable for protein O-GlcNAcylation which is a post-translational protein modification that is involved in various signaling, immune, and structural functions (Akella et al., 2019). O-GlcNAcylation modulates cancer metabolism by regulating glycolysis via HIF-1 $\alpha$  and GLUT1 (Ferrer et al., 2014), controlling flux into the PPP via PFK1 (Yi et al., 2012), influencing glutamine uptake (Wellen et al., 2010), and by modifying the bioenergetic sensor AMPK (Bullen et al., 2014). O-GlcNAc transferase (OGT) is responsible for the covalent O-GlcNAcylation modification of proteins by UDP-GlcNAc and is related to DNA damage responses in TNBC (Barkovskaya et al., 2019).

A substantial number of studies have confirmed the connection between the HBP and EMT (Akella et al., 2019; K. Taparra et al., 2016). EMT increases the UDP-GlcNAc level and modulates the O-glycan and N-glycan compositions, and several EMT markers—for example, E-cadherin, TGF- $\beta$ R, and EGFR—are modified by N-linked glycosylation (X. Li et al., 2016). The EMT-TF Snail can be stabilized by O-GlcNAcylation modification to instigate the EMT program, which directly links glucose metabolism to EMT (S. Y. Park et al., 2010). In addition, the EMT-associated regulators TGF- $\beta$ , NF- $\kappa$ B, and FOXO-1 are all subject to potential O-GlcNAcylation modification (Akella et al., 2019). Given that the HBP resides at the crossroads of many central metabolic pathways, and the downstream product, UDP-GlcNAc, is primarily responsible for protein O-GlcNAcylation and O-linked and N-linked glycans that regulate many oncogenes, targeting the HBP is appealing for cancer therapeutics (Akella et al., 2019; Chiaradonna et al., 2018).

UDP-GlcNAc, along with UDP-glucuronic acid (UDP-GlcA), UDP-xylose, UDP-galactose, and UDP-N-acetylgalactosamine, are UDP-activated nucleotide sugars serving as building blocks for the ubiquitous glycosaminoglycans (GAGs). GAGs are polysaccharides that comprise four main categories, namely chondroitin sulfate/dermatan sulfate (CS/DS), heparin/heparan sulfate (HS), hyaluronan (HA), and keratan sulfate (KS). GAGs partake in a plethora of cellular functions spanning signal regulation, ECM remodeling, TME modulation, and immunosurveillance, which are culpable for oncogenesis (Morla, 2019). For example, HA is abundant in the ECM and can bind to and activate CD44 (C. Chen et al., 2018). HS proteoglycans are involved in the ligand-receptor complex formation in FGF-2 signaling in breast carcinomas (Mundhenke et al., 2002). The CS

proteoglycan versican is highly deposited in the peritumoral stroma in primary breast cancer and may promote cancer progression (Ricciardelli et al., 2002).

***UGDH: One of the main enzymes in glycan metabolism identified in this project***

In 1956, Strominger and colleagues identified and purified the enzyme UGDH from calf liver homogenate and confirmed that UGDH catalyzes UDP-glucose (UDP-Glc) to UDP-GlcA (Strominger et al., 1957). UDP-GlcA is an indispensable unit of GAGs (such as HA) and proteoglycans and a glucuronyl donor to endogenous and exogenous compounds conjugated as glucuronides in liver. UGDH uses NAD<sup>+</sup> as an oxidant and catalyzes UDP-Glc to UDP-GlcA while generating NADH in a two-step oxidation.

Earlier studies had investigated the functions of UGDH in embryonic development. UGDH shapes tissue morphogenesis via HA-mediated pressure stimulus (Munjal et al., 2021). Concerning the functions of UGDH in glucuronidation reaction of xenobiotics, UGDH was reported to be under the influence of xenobiotic treatments (Vatsyayan et al., 2005). HA is one of the main products produced by UGDH and constitutes the main part of the ECM. Thus far, the pathological effects of UGDH have mainly been attributed to the function of UGDH in the generation of UDP-GlcA and the subsequent production of HA and other GAGs and proteoglycans. HA may facilitate cancer invasion, chondrogenic matrix accumulation, and osteoarthritis (Clarkin et al., 2011; Clarkin et al., 2011; Passi et al., 2019; Wen et al., 2014). Given the significance of the ECM and HA composition in cancer progression (Vigetti and Passi, 2014) and the essential role of UGDH in HA formation, the biological function of UGDH in cancer development has been increasingly appreciated in the last five years.

*UGDH in cancer*

UGDH inhibition can decrease the proliferation rate of the human breast cancer cell line MCF-7 and the HCT-8 colorectal cell line (Hwang et al., 2008; T. P. Wang et al., 2010) and reduce the prostate cancer cell colony formation (Scoglio et al., 2016). Moreover, small interfering RNA (siRNA)-mediated UGDH knockdown in these cells delayed the spheroid formation and hampered the cell migration and invasion abilities, which were compensated by adding exogenous HA (T. P. Wang et al., 2010). In addition to prostate cancer, breast cancer, and colorectal cancer, UGDH in cancer aggressiveness and chemoresistance has also been implicated in melanoma, non-small cell lung carcinoma, glioblastoma, and ovarian cancer (Deen et al.,

2016; Lin et al., 2020; Oyinlade et al., 2018; Paul et al., 2016). UGDH knockdown decreased cell proliferation and migration in glioblastoma cells *in vitro* and inhibited tumor growth and migration *in vivo* (Oyinlade et al., 2018). Similarly, UGDH was highly expressed in a very invasive ovarian cell line, and UGDH knockdown impaired ovarian cancer cell proliferation, migration, and ovarian tumor growth in xenograft mouse models (Lin et al., 2020). UGDH is negatively correlated with the survival of patients with TNBC who received chemotherapy, perhaps due to the glucuronidation of xenobiotics by UGDH (Vitale et al., 2021). The inhibitory effects of the UGDH inhibitor—4-methylumbelliferone (4-MU)—on breast cancer formation, growth, and invasion via inhibiting HA synthesis manifest similarly to those of UGDH knockouts (Arnold et al., 2019).

Recently, Wang et al. (2019) reported that phosphorylation of UGDH by EGFR at tyrosine 473 could enhance the stability of SNAI1 mRNA to facilitate lung cancer metastasis by attenuating the UDP-Glc-mediated SNAI1 mRNA decay. This study has proposed a new mechanism of UGDH in cancer progression mediated by UDP-Glc (X. Wang et al., 2019). Finally, UGDH could serve as a sero-diagnostic marker to facilitate the prognosis of patients with early lung adenocarcinoma (Hagiuda et al., 2019).

#### *UGDH in EMT*

The first evidence of UGDH in EMT was the observation of upregulation of UGDH in epithelial breast cancer cells after *E*-cadherin knockdown (Vergara et al., 2015). Later, Arnold and colleagues noticed that UGDH was one of the top dysregulated genes in mesenchymal-like breast cancer, and several mesenchymal-like characteristics could be inhibited by depleting UDP-GlcA (Arnold et al., 2019). Furthermore, UGDH depletion inhibited the EMT-TFs SNAI1 and Smad interacting protein-1 (SIP-1) and the mesenchymal marker *N*-cadherin but increased the epithelial marker *E*-cadherin (Lin et al., 2020; X. Wang et al., 2019). Even though Teoh et al. (2020) had observed similar effects of UGDH on UDP-GlcA production, breast cancer migration, and *in vivo* tumor growth and metastasis, they claimed that UGDH knockout did not impair EMT in the mouse mammary cancer cells on account of the increased EMT markers fibronectin (FN1) and EMT-TFs homeobox protein SIX1. Nevertheless, they also reported *E*-cadherin upregulation with UGDH knockout (Teoh et al., 2020).

*UGDH regulation and the regulatory roles of UGDH*

UGDH can be upregulated by TGF- $\beta$  and downregulated by hypoxia and is regulated by the transcription factor specificity protein 1 (SP1) (Bontemps et al., 2003). In addition, the zinc-finger transcription factor c-Krox, in cooperation with SP1/SP3, inhibits UGDH in chondrocytes, while the steroid hormone estradiol 17 $\beta$ -E<sub>2</sub> stimulates UGDH in articular chondrocytes (Beauchef et al., 2005). Furthermore, the Epstein-Barr virus oncogene latent membrane protein 2A (LMP2A) can upregulate UGDH by activating the extracellular signal-regulated kinase (ERK) and PI3K/AKT pathways and eventually initiate the binding ability of SP1 to the promoter region of UGDH (Pan et al., 2008). The regulation of MEK-ERK on UGDH was confirmed in articular surface cells; however, the TGF- $\beta$ -mediated p38<sup>MAPK</sup> activity is more dominant in regulating UGDH in these cells (Clarkin et al., 2011). The regulatory effects of transcription factors SP1/3 and c-Krox on UGDH and the involvement of p38<sup>MAPK</sup> were validated in human primary chondrocytes (Wen et al., 2014). UGDH is also a downstream target of the binding activity of Krüppel-like factor 4 (KLF4) and methylated CpG and is regulated by KLF4 in a DNA methylation-dependent manner (Oyinlade et al., 2018). A recent study showed that the glycoprotein Slit2 has regulatory effects on UGDH in CD34+ fibrocytes isolated from peripheral blood mononuclear cells (Fernando and Smith, 2021). Taken together, UGDH is regulated by multiple signaling pathways at the genetic, post-transcriptional, and post-translational levels.

Interestingly, studies also suggest that UGDH can exert regulatory effects, affecting the expression of important cellular regulators including Notch1 and peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ) and, subsequently, lipid metabolism (Arnold et al., 2019; Zimmer et al., 2016). In addition, the regulatory role of UGDH has also been implicated in controlling the cell cycle checkpoints via p21 and p27 and affecting the ERK/MAPK pathway and phosphorylation of AKT (Hagiuda et al., 2019; Lin et al., 2020). Lastly, UGDH catalyzes a reaction that can significantly influence the NAD<sup>+</sup>/NADH ratio, and NAD<sup>+</sup> and NADH strongly control the metabolic activity of SIRT1 (Anderson et al., 2017).

***GFPT2: The second main glycan enzyme identified in this project***

In 1999, human GFPT2 was first subcloned by polymerase chain reaction (PCR), based on the expressed-sequence tag (EST) sequence that is similar to human GFPT1. The homology between human GFPT1 and human GFPT2 is 75%-80%, while there is 97%-98% similarity between human GFPT2 and mouse GFPT2 (Oki et al., 1999). Structure analysis reveals that the

glutaminase domain of GFPT2 possesses a flexible loop structure, potentially contributing to its functional flexibility. Enzyme kinetics suggest that the efficiency of GFPT2 in synthesizing glucosamine-6-phosphate is lower than other GFPTs. GFPT2 also hinders the productivity of glutamine hydrolysis (Oliveira et al., 2021).

Although GFPT2 has enzymatic activities that resemble GFPT1, these two enzymes have different tissue distributions (Hu et al., 2004) and are regulated differently (Hu et al., 2004; Kuang et al., 2008; Ruegenberg et al., 2020). GFPT1 is negatively regulated by its product, UDP-GlcNAc, while GFPT2 is only weakly affected (Hu et al., 2004). GFPT2 reacts to EGF stimulation and is upregulated in the presence of EGF in mice (Richani et al., 2014). GFPT2 has been identified as a molecular marker for embryonic definitive endoderm (Lawton et al., 2013; P. Wang et al., 2012). Many studies have suggested GFPT2 has a distinct gene expression pattern associated with ethnicity and is involved in antipsychotic-medication-induced weight gain, excess adiposity, diabetes, chronic obstructive pulmonary disease, and cardiac/cardiovascular functions (Belke, 2011; Coomer and Essop, 2014; Gabel et al., 2017; Kresovich et al., 2017; Prasad et al., 2010; H. Yu et al., 2016; H. Yu et al., 2021; H. Zhang et al., 2004). Recent studies show that GFPT2 is involved in myocardial infarction in mice. GFPT2 plays a role in the hypoxia/reoxygenation-induced myocardial cell damage where GFPT2 is regulated by KLF5-miR-27a axis and activates the TGF- $\beta$ /Smad2/3 signaling pathway (Tian et al., 2021). GFPT2 expression is not presented in cardiac myocytes but is highly abundant in cardiac fibroblasts (Nabeebaccus et al., 2021). It serves as a marker for a subpopulation of cardiac fibroblasts and affects the functions of these cells (Li Wang et al., 2021).

#### *GFPT2 regulation in cancer and EMT*

GFPT2 is reported to be under the regulation of bone morphogenetic protein 2 (BMP-2), NF- $\kappa$ B, SIRT6, inositol-requiring enzyme (IRE1 $\alpha$ ), spliced X-Box binding protein 1 (sXBP1), and FoxO1 (Al-Mukh et al., 2020; Mirmalek-Sani et al., 2009; Panarsky et al., 2020; Qiao et al., 2021; Szymura et al., 2019). It regulates NF- $\kappa$ B (p65) via O-GlcNAcylation to form a feedback loop for its regulation (L. Liu et al., 2020). GFPT2 correlates with SNAI1 and TWIST1 (Kekoa Taparra et al., 2019), is associated with tumor progression, and is upregulated after EMT in breast cancer, non-small cell lung cancer (NSCLC), lung adenocarcinoma, colon adenocarcinoma, and serous ovarian cancer (J. Kim et al., 2020; D. Li et al., 2021; Shaul et al., 2014; Simpson et al., 2012; Szymura et al., 2019; Kekoa Taparra et al., 2019; Verbovsek et al., 2014; L.

Zhang et al., 2021; W. Zhang et al., 2018; Zhou et al., 2019). It interacts with TGF- $\beta$ 1 and increases TGF- $\beta$ 1 production (Prasad et al., 2010). In turn, TGF- $\beta$ 1 upregulates GFPT2 gene expression (W. Zhang et al., 2018). GFPT2 can promote EMT in serous ovarian cancer via O-GlcNAcylation of  $\beta$ -catenin (Zhou et al., 2019). It holds potential as a therapeutic target in NSCLC with concurrent KRAS and LKB1 mutations that emulates claudin-low breast cancer (H. S. Kim et al., 2013; J. Kim et al., 2020), indicating the role of the LKB1-AMPK pathway in GFPT2 regulation. GFPT2 is one of the upregulated claudin-low signature genes in TNBC (H. S. Kim et al., 2013; Prat et al., 2010). Its level is elevated in breast cancer biopsies (Oikari et al., 2018).

#### *The uncanonical functions of GFPT2*

Although most studies on GFPT2 have focused on the function of GFPT2 in generating UDP-GlcNAc and the downstream effects of O-GlcNAcylation on specific targets, the regulatory role of GFPT2 is not limited to protein O-GlcNAcylation. GFPT2—highly expressed in CAFs in lung adenocarcinoma—correlates with glucose uptake in the TME and facilitates tumor progression via metabolic reprogramming in TME (W. Zhang et al., 2018). Studies have reported the role of GFPT2 in responding to oxidative stress (Nivet et al., 2013; Zitzler et al., 2004). GFPT2 protects cells from peroxide-induced oxidative stress (Zitzler et al., 2004). Homozygosity of one GFPT2 mutation leads to increased ROS in spermatozoa and decreases sperm mobility in men (Askari et al., 2019). A recent study showed that the mRNA level of GFPT2 decreased after glyoxalase 1 (GLO1) knockout, and GLO1 detoxifies methylglyoxal in a glutathione-dependent manner (Jandova and Wondrak, 2020). It is worth noting that GFPT2 expression changes upon virus infections (X. Xu et al., 2020; Y. Zhao et al., 2019), and treating cells with scrambled siRNAs could increase GFPT2 expression (Oikari et al., 2016). Recent studies have reported that GFPT2-overexpressing tumor cells are associated with chemoresistance, TME regulation and interference with immune cells in immunotherapy (X. Ding et al., 2022; J. Li et al., 2022).

#### **1.3.5 Metabolic reprogramming of one-carbon metabolism and the transsulfuration pathway**

One-carbon metabolism includes biochemical pathways encompassing both the folate and methionine cycles to produce and transfer one-carbon units or methyl groups. Serine, glycine, and choline are all active one-carbon unit donors (Newman and Maddocks, 2017). The folate cycle plays a part in nucleotide synthesis and antioxidant defense, and the methionine cycle

regulates the provision of methyl groups for DNA, histone, and protein methylation modifications, which are major epigenetic regulations in mammalian cells (Friso et al., 2017; M. Yang and Vousden, 2016). The ratio of two intermediates in the methionine cycle—S-adenosylhomocysteine and S-adenosylmethionine (SAH:SAM)—is reflective of the cellular methylation capacity (Weber and Birsoy, 2019). Cancer cells may utilize one-carbon units for nucleotide synthesis, methylation modification, and NADH/NADPH production. To complete the methionine cycle, methionine needs to be recycled by re-methylation of homocysteine which participates in the transsulfuration pathway for synthesis of cysteine and glutathione.

In the transsulfuration pathway, homocysteine is converted into cystathionine and further into cysteine on which tumor cells are heavily reliant (J. K. M. Lim et al., 2019). The end-product, cysteine, is a precursor for *de novo* glutathione synthesis where glutamate and glycine are also involved, and glutathione is a ubiquitous antioxidant. Zhu and colleagues reported that with limited sources of extracellular cysteine, tumor cells rely on the transsulfuration pathway for *de novo* cysteine synthesis to support cell growth (J. Zhu et al., 2019). Apart from the production of cysteine, the transsulfuration pathway is also primarily responsible for the endogenous production of the gaseous signaling molecule hydrogen sulfide (H<sub>2</sub>S). Weber and colleagues argued that H<sub>2</sub>S, cystathionine, and homocysteine in the transsulfuration pathway are all possibly accountable for increased tumor growth (Weber and Birsoy, 2019). Cystathionine-β-synthase (CBS) is the first and rate-limiting enzyme in the transsulfuration pathway. It catalyzes the conversion of homocysteine into cystathionine, the upregulation of which induces altered expression of genes in various pathways that favor carcinogenesis and EMT. The CBS-H<sub>2</sub>S signaling axis promotes cell growth and metastasis in colorectal cancer (Phillips et al., 2017). Cystathionine is recognized as an oncometabolite in breast cancer, protects cancer cells against ROS and drug-induced apoptosis, and maintains homeostasis of both mitochondria and ER (Sen et al., 2016).

### **1.3.6 Metabolic reprogramming in ROS regulation and mitochondrial metabolism**

Excess ROS can be detrimental for cancer cells. The superoxide anion (O<sub>2</sub><sup>-</sup>) generated from the ETC is one of the main sources of ROS (Idelchik et al., 2017). The input of glutamine has strong impacts on ROS via glutamate, a precursor for *de novo* glutathione synthesis. Cancer cells have increased the defense mechanism against ROS; it can be induced by matrix detachment



during cancer cell circulation (L. Jiang et al., 2016). Detached and disseminated cancer cells in circulation form clusters to increase their antioxidant defense (Labuschagne et al., 2019). ROS activates a variety of EMT-TFs and ultimately facilitates EMT and the progression of tumorigenesis (Hayes et al., 2020).

The loss of anchorage dependency of cancer cells causes mitochondrial perturbations and ROS production, which can be ameliorated by cell clustering to achieve successful distant colonization (Labuschagne et al., 2019). Oxidative phosphorylation (OXPHOS) and mitochondrial functionality are most affected in cancer and associated with EMT (Edoardo Gaude and Frezza, 2016). Aside from ROS formation, mitochondrial dysregulation also results in the accumulation of fumarate, succinate, and 2-hydroxyglutarate (2-HG), which have been shown to be oncogenic and causes functional deficits in mitochondrial outer membrane permeabilization (MOMP) to protect neoplastic cells against regulated cell death (RCD) (E. Gaude and Frezza, 2014; Izzo et al., 2016). Metabolic dysregulation in mitochondria affects malignant transformation, cancer cell proliferation, resistance to RCD, interaction with the stroma, metastatic dissemination, resistance to therapeutics, immunosurveillance, and EMT (Porporato et al., 2018). The mitochondrial metabolite fumarate can repress the demethylation of antimetastatic miR-200 mediated by the ten-eleven translocation (TET) methylcytosine dioxygenases to suppress miR-200 and overexpress EMT-TF ZEB2 (Sciacovelli et al., 2016).

Mitochondrial dysfunction affects H<sub>2</sub>S homeostasis, and H<sub>2</sub>S is associated with ROS production and oxidative stress (Quinzii and Lopez, 2021). Sulfide quinone oxidoreductase (SQOR) resides at the mitochondrial membrane and catalyzes the commitment step of H<sub>2</sub>S oxidation by coupling the reduction of coenzyme Q (CoQ, also referred to as ubiquinone), which is a component of the mitochondrial ETC. The H<sub>2</sub>S level is closely regulated by both CoQ and glutathione (Quinzii and Lopez, 2021) and tightly controlled by SQOR via H<sub>2</sub>S oxidation (Jackson et al., 2012).

### **1.3.7 GPC and NAA**

#### ***GPC is a potential target for breast cancer treatment***

Activated choline metabolism is a hallmark of cancer progression (Glunde et al., 2011; Egidio Iorio et al., 2016). GPC is involved in the catabolic pathway of phosphatidylcholine (PtdCho), which is the most abundant phospholipid in the cell membrane to produce choline. GPC is vital for choline synthesis, and

most studies have focused on the Kennedy pathway and the production of PtdCho in choline metabolism. Recently, however, an abnormal GPC level in cancer is emerging as a target for cancer treatment (Glunde et al., 2011; Sonkar et al., 2019). In addition, GPC exerts cellular osmotic regulatory functions under osmoregulation of the phospholipases and the glycerophosphocholine phosphodiesterases (GPC-PD) GDPD5, which was first found in renal medullary cells (Gallazzini and Burg, 2009). GDPD5 knockdown results in the accumulation of GPC and the increased ability of cells to produce proteoglycans, which indicates that GDPD5 regulates the intracellular osmotic stress via GPC (Okazaki et al., 2019). Silencing GPC-PD decreases the level of lipid metabolites, suggesting a close relationship between GPC and downstream lipid metabolism (Stewart et al., 2012).

Breast cancer cells treated with drugs (the PI3K inhibitor BEZ235, the Hsp190 inhibitor 17-AAG, doxorubicin, the nonsteroidal anti-inflammatory agent indomethacin, and the AKT inhibitor perifosine) exhibit increased GPC via upregulation of phospholipase A2 (PLA2) (Brandes et al., 2010; Cheng et al., 2017; Glunde et al., 2006; Siver A. Moestue et al., 2013; Su et al., 2012), and GPC possesses antioxidant and anti-inflammatory effects (Tokes et al., 2015). Even though the treatment of cells in culture with chemical reagents results in increased GPC expression, a decrease in GPC indicates better survival in patients during neoadjuvant chemotherapy (M. D. Cao et al., 2012). Earlier studies had used the phosphocholine (PC)/GPC ratio as an indicator for tumor malignancy, an approach that has aroused inconsistency and controversy. The elevated PC/GPC ratio, or the so-called GPC-to-PC switch, has been associated with escalated malignant transformation *in vitro* (Cheng et al., 2017). However, there is a lower PC/GPC ratio in the most aggressive animal tumor models and patients with the worst survival (M. D. Cao et al., 2016; Siver A. Moestue et al., 2010). In both basal-like breast cancer xenografts and tissue samples from patients with TNBC, GPC concentrations are higher with a GPC/PC ratio > 1 compared with luminal-like and ESR<sup>+</sup>/PR<sup>+</sup> tumors (Siver A. Moestue et al., 2010). These findings indicate that the GPC level is positively correlated with breast tumor malignancy (S. A. Moestue et al., 2012). GPC plays important roles in choline metabolism in cancer, but the regulatory mechanism of GPC is poorly understood. Both choline kinase alpha (CHK $\alpha$ ) and GDPD5 have been positively correlated with the GPC concentration in breast cancer xenografts (Grinde et al., 2014). PI3K, HIF-1 $\alpha$ , and HIF-2 $\alpha$  have been reported to regulate the GPC level (Bharti et al., 2018; Lau et al., 2017).

### ***NAA in cancer***

NAA is synthesized by aspartate *N*-acetyltransferase (NAT8L) from acetyl-CoA and aspartate and metabolized by aspartoacylase (ASPA) to generate aspartate and acetate, which is further metabolized into acetyl-CoA for energy metabolism and lipid synthesis. NAA is the second most abundant metabolite in the brain, and most studies on NAA are related to brain pathology (Bogner-Strauss, 2017; Miyake et al., 1981). NAA is higher in tumors than normal tissues, a phenomenon that might be caused by increased expression of NAT8L and associated with glutamine availability (Lou et al., 2016; Zand et al., 2016). Several studies have reported that overexpression of NAT8L facilitates cancer progression, and silencing NAT8L can decrease the NAA level, inhibit cell proliferation, and reduce tumor growth (Lou et al., 2016; Zand et al., 2016). NAT8L is associated with worse survival in patients with different cancer types and under the control of oncogene RhoC (Wynn et al., 2016). Conversely, ASPA is downregulated in cancer, suggesting that tumor cells rely on NAA rather than its downstream products. The secreted NAA in the peripheral blood may serve as a clinical biomarker for cancer malignancy. The exact roles of NAA, NAT8L, and ASPA in cancer metabolism need to be further elucidated (Bogner-Strauss, 2017).

### ***GPC and NAA are connected via lipid metabolism***

Choline—including GPC—and NAA are two of the metabolites well-detected in the brain via magnetic resonance spectroscopy (MRS). Tsougos and colleagues reported that the peritumoral choline/NAA ratio could differentiate glioblastomas from intracranial metastasis with high specificity (Tsougos et al., 2012). GPC and NAA are not components of the same metabolic pathway. Nevertheless, both GPC and NAA are closely connected to lipid metabolism that is partially regulated by the potent regulator PPAR $\gamma$ , a transcription factor involved in a variety of metabolic activities (Ahmadian et al., 2013). PPAR $\gamma$  activation may contribute to cancer cell proliferation, apoptosis, angiogenesis, and metastasis (Yousefnia et al., 2018). The exact relationship between GPC and NAA is elusive.

## **1.4 The D492, HMLE, and PMC42 EMT cell models**

Different cells of origin, or cancer-initiating cells, may reflect and contribute to heterogeneous behaviors and phenotypes in cancer. Each cell line reflects tumors *in vivo* differently and has different clinical relevancy. A better understanding of the cell lines chosen to study is a prerequisite to drawing

reliable conclusions from the investigation. In the next section, I will briefly review the cell lines used in this project.

#### **1.4.1 The isogenic EMT cell lines D492, D492M, and D492HER2**

In 2002, the D492 cell line was generated in the pursuit of finding neoplastic stem cells in the human mammary gland on the assumption that breast cancer cells are of a luminal epithelial lineage and originate from terminal duct lobular units (TDLU). Patients underwent reduction mammoplasty for cosmetic reasons, and a group of MUC-/ESA<sup>+</sup> cells were selected from primary cultures of the biopsies. These cells were further immortalized with human papilloma virus (HPV)-16 E6/E7 to generate the immortal D492 cell line. These cells are of a luminal epithelial lineage, but they express both luminal (K8, K19) and myoepithelial (K5/6, K14) cytokeratins, thus showing the basal-like phenotype. D492 has epithelial stem cell properties, and it can differentiate into luminal and myoepithelial cells and form the entire TDLU with branching morphogenesis. D492 cells are non-tumorigenic and have a diploid karyotype (Gudjonsson et al., 2002).

Nine years after the generation of D492, the D492M cell line was spontaneously generated by co-culturing D492 cells with human breast endothelial cells (BRENCs) inside a reconstituted basement membrane (rBM). By isolating and subculturing a single spindle-like mesenchymal cell colony, the D492M cell line was established and acquired a stable passaging ability. Compared with the parent cell line D492, the D492M cell line has downregulated epithelial *E*-cadherin as well as keratins 5, 6, 14, and 19, and upregulated mesenchymal vimentin, *N*-cadherin, alpha-smooth muscle actin ( $\alpha$ -SMA), Thy-1, thrombin receptor (PAR1), and CD70. In addition, the EMT-TFs FOXC1 and FOXC2 are upregulated in D492M. In addition to a gain of mesenchymal properties, D492M cells possess a cancer stem cell phenotype, supported by the increased proportion of the CD44<sup>high</sup>/CD24<sup>low</sup> cells. Like D492, the D492M cell line is not tumorigenic. The EMT-inducing ability of the endothelial cells is related to the endothelium-derived soluble factors. The exact factors that induce EMT are elusive; however, hepatocyte growth factor (HGF) has been confirmed to play a role in the formation of D492M cells because HGF inhibition could significantly decrease the number of D492M colonies (Sigurdsson et al., 2011). After establishing the mesenchymal D492M cell line, efforts have been made to reverse the EMT process in D492M. Co-expression of miR-200c-141 and  $\Delta$ Np63 restored the properties of the parent cell line D492 (Hilmarsdóttir et al., 2015).

EMT in the D492M cell line was induced by spontaneous stimulation, while the mesenchymal features in the D492HER2 cell line were introduced by forced expression of the oncogene HER2 (ERBB2) in D492. The D492HER2 cells have increased EMT-TF ZEB1 and have lost epithelial K14 and K19, *E*-cadherin, *P*-cadherin, and p63. Furthermore, miR-200c-141, the epithelial morphogenic regulator, is also downregulated in D492HER2 cells. Apart from the EMT markers, D492HER2 cells have also gained the loss-of-contact growth ability. Based on the observation that HER2 overexpression decreased the EGFR level in D492, overexpressing EGFR in D492HER2 cells restored some of the epithelial markers, including K14, K19, and *E*-cadherin. However, this partial reverse of EMT in D492HER2 cells by EGFR overexpression only appeared in 3D rBM culture. Furthermore, injecting D492HER2 cells into mammary fat pads of NSG<sup>TM</sup> mice gave rise to tumors, which were attenuated by EGFR overexpression (Ingthorsson et al., 2016). The D492HER2 cell line appears to represent a more intermediate state of EMT compared with the D492M cell line (Morera et al., 2019).

#### **1.4.2 The common EMT cell model HMLE-HMLEM**

A widely studied EMT cell model is the HMLE cell lines, which were established in Professor Weinberg's lab in 2000. Like D492, the HMLE epithelial cell line was derived from normal breast biopsies of reduction mammoplasty. They express *E*-cadherin, basal K14, and luminal K18, indicating these cells are bipotential or multipotential. To immortalize these cells, both the simian virus 40 (SV-40) large T antigen (LT) oncogene and the hTERT gene that encodes the catalytic subunit of the human telomerase enzyme were introduced. The immortalized cells lack the ability to form tumors in nude mice (Elenbaas et al., 2001).

The generation of the mesenchymal counterpart of the HMLE cell line, referred to as HMLEM in this project, was based on two assumptions: First, CSCs have acquired EMT, and to identify targets selectively aiming at CSCs and leaving out the non-CSCs, EMT needs to be induced and the key regulatory genes of mesenchymal states need to be studied. Second, forced overexpression of specific EMT inducers does not adequately reflect the behaviors of the mesenchymal cells that arise *in vivo*. Hence, the HMLEM cell line, referred to as naturally arising mesenchymal cells (NAMECs) in the original publication, was produced by screening and isolating cells with stable mesenchymal phenotypes after 1-minute trypsinization and replating the HMLE cells. These cells have elevated levels of the mesenchymal markers vimentin, *N*-cadherin, and FN1 as well as the EMT-TFs TWIST, Snail, Slug,

and ZEB1. They have lost *E*-cadherin. Like the D492M cell line, HMLEM cells are predominately CD44<sup>high</sup>/CD24<sup>low</sup>. HMLE cells, deemed non-CSCs, possess stronger EGFR signaling than HMLEM cells, which are considered to be CSCs, while HMLEM cells have switched to PDGFR signaling. This agrees with the observation in D492HER2 that HER2 overexpression decreased the EGFR level in D492 cells, and EGFR overexpression impeded the mesenchymal characteristics in D492HER2 cells (Ingthorsson et al., 2016; Tam et al., 2013).

### **1.4.3 The MET cell model PMC42LA-PMC42ET**

The third EMT cell model used in this project is the PMC42 cell lines, comprising the mesenchymal PMC42ET cell line from Dr. Robert Whitehead (Whitehead et al., 1983) and the epithelial PMC42LA cell line generated by Dr. Leigh Ackland (Ackland et al., 2001). There are two major differences between the PMC42 model and the other two EMT models. First, the mesenchymal PMC42ET cells were isolated from a pleural effusion from a woman with metastatic breast cancer without the need for immortalization. Second, the epithelial PMC42LA was established from its mesenchymal counterpart by initiating MET. These cells possess karyotypic heterogeneity and chromosomal instability. The mesenchymal PMC42ET cells are heterogenous and believed to have stem cell characteristics on account of their ability to differentiate into several morphological phenotypes. They do not form tumors *in vivo*. A hormone mixture comprising estrogen, progesterone, dexamethasone, insulin, and prolactin, in combination with a porous filter coated with Matrigel, induced cell transformation in PMC42ET cells to generate PMC42LA cells. The PMC42LA cell line is a stable variant of the PMC42ET cell line treated with lactogenic hormones and can be stimulated by EGF (Ackland et al., 2003) and secreted factors from carcinoma-associated fibroblasts (Lebret et al., 2007) to undergo EMT. PMC42LA cells express the milk-specific proteins  $\beta$ -casein, K8 and K18, and *E*-cadherin (H. Hugo et al., 2007).

## **1.5 The multi-omics era in molecular biology**

The insufficiency of the genomic and transcriptomic data to reflect and predict the actual protein expression and the cellular proteoform acquires views directly from the proteomic perspective (Akbani et al., 2014). Studies have confirmed the better performance of proteomic profiling for clinical relevance compared with transcriptomic approaches (Akbani et al., 2014; J. Wang et al., 2017). This project was built on proteomic analysis of the EMT cell lines.

Hereafter, I will review proteomics and metabolomics in the multi-omics era propelled by myriad technical achievements.

### **1.5.1 The rise of proteomics**

Proteomics is the large-scale systematic study of a collection of proteins and protein post-translational modification (PTMs) with the aid of the mass-spectrometry-based detection methods and computational analysis for protein identification and quantification to generate biological insights. It has been utilized extensively in EMT and TNBC research (Mathias and Simpson, 2009; Miah et al., 2016). Apart from the global proteome, there are proteomic approaches to study specific sub-sections of the proteome, such as the membrane proteome, the glycoproteome, the secretome, lysosomal proteins, and palmitoylation proteomics, and the most frequently studied phosphoproteomics (Buccitelli and Selbach, 2020).

#### ***Bottom-up, middle-down, and top-down proteomic approaches***

The idea of analyzing the whole proteome arose in 1970s with the development of two-dimensional (2D) gel electrophoresis (James, 1997). As mass spectrometry (MS) has improved, especially the introduction of the electrospray ionization (ESI) technique in 1989, the situation for proteomics has advanced tremendously (Fenn et al., 1989; Mann, 2016). With the advancement of liquid chromatography (LC), the advent of bioinformatics, the foundation of online databases to facilitate the downstream proteomic analysis, and the breakthroughs in the electrospray nano-LC-MS/MS technique, MS-based proteomics has improved dramatically in terms of sensitivity, complexity, and throughput (Mann, 2016).

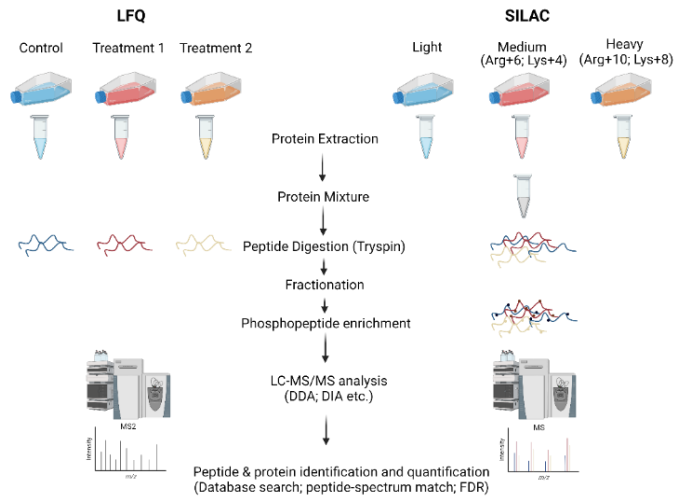
There are two alternative strategies in proteomics, namely top-down proteomics, where intact protein ions or large protein fragments are subjected to LC-MS, and bottom-up proteomics, which analyzes digested peptides and is the most widely used proteomic approach (Aebersold and Mann, 2016). In bottom-up proteomics, proteins are extracted from the biological samples followed by enzymatic digestion into peptides by, for example, trypsin, and the resulting mixture of peptides (typically contain 7-20 amino acid residues) are separated, ionized, and analyzed via LC-MS/MS. The raw data generated from the LC-MS/MS platform are analyzed by specifically designed computational pipelines and further explored for biological meanings. Top-down proteomics is still in its early stage of development to improve the depth of discovery, and bottom-up proteomics holds several pitfalls including discovery of PTMs; this limitation hinders the further growth of the proteomic

field. To overcome the disadvantages of both strategies, scientists have recently proposed a middle-down approach. Similarly to bottom-up proteomics, the middle-down approach relies on proteases, such as OmpT, Sap9, and IdeS, to generate longer peptides. Nevertheless, bottom-up proteomics remains the workhorse of proteomic sequencing (Pandewari and Sabareesh, 2019).

### ***Discovery and targeted proteomics***

Discovery (or “shotgun”) proteomics by means of data-dependent acquisition (DDA), targeted proteomics by selected reaction monitoring (SRM), and multiplexed fragmentation of all peptides via data-independent acquisition (DIA) are three main approaches in bottom-up proteomics. Antibody-based immunoassays, for example, western blot and reverse-phase protein array (RPPA), are widely utilized to target specific proteins, relying on valid antibodies that are expensive and not available for every protein. The drawbacks of the immunoassays can be overcome by targeted proteomic approaches, which would be groundbreaking for cancer research and help identify clinically significant molecular biomarkers (Faria et al., 2017). Several studies have implemented combining discovery and targeted proteomics and have shown promising results (Biarc et al., 2014; Hill et al., 2009). That having been said, “shotgun” proteomics with DDA is still the most frequently used method in the proteomic community and was carried out in this project for the systematic proteomic profiling of the EMT cell lines. Two independent DDA proteomic experiments based on two disparate quantitative strategies were conducted in this project to ensure data accuracy and repeatability, and potential targets were selected based on a deep understanding of the EMT cell lines and a scrupulous literature review (**Figure 3**).





**Figure 3.** General workflow of LFQ and SILAC proteomics.

D492, D492M, and D492HER2 cells were subjected to LFQ and SILAC proteomic analysis in this project. LFQ: label-free quantification; SILAC: stable isotope labeling by amino acids in cell culture; Arg: arginine; Lys: lysine; DDA: data-dependent analysis; DIA: data-independent analysis; LC: liquid chromatography; MS: mass spectrometry; FDR: false discovery rate.

### **Quantitative proteomics**

Quantitative proteomics can be categorized into relative and absolute quantification, as well as labeling and label-free quantitative methods. Label-free quantification (LFQ) and intensity-based absolute quantification (iBAQ) are common relative and absolute proteomic quantification algorithms, respectively (Cox et al., 2014; Nagaraj et al., 2011). Stable isotope labeling by amino acids in cell culture (SILAC), stable isotope dimethyl labeling, isobaric tags for relative and absolute quantitation (iTRAQ), and tandem mass tag (TMT) labeling are popular labeling methods applied in proteomics. The SILAC approach is used to cultivate cells in different types of media containing either normal amino acids (arginine and lysine) or amino acids labeled with heavy isotopes ( $^{13}\text{C}$ ,  $^{15}\text{N}$ , or deuterium). Cells fully incorporated with the respective “light” or “heavy” labels are mixed and analyzed simultaneously in LC-MS/MS to generate ratios for all the proteins while reducing batch effects (Ong and Mann, 2006). In this project, we mainly employed two proteomic approaches established on two quantification methods, namely LFQ and SILAC, supported by the absolutely quantitative iBAQ. Both approaches have been widely used in EMT research (**Figure 3**) (Biarç et al., 2014; D. Chen et al., 2015; Palma Cde et al., 2016; G. R. Yan et al., 2011).

### ***Secretomics and phosphoproteomics***

Eukaryotic cells are highly compartmentalized to partition cellular functions, an organization that accentuates the importance of protein subcellular localization and dynamics. The protein composition of different organelles, along with the secreted protein mediators (secretome), can be enriched, purified, and analyzed with proteomic approaches. By analyzing the secreted soluble molecules in the extracellular compartment, the factors associated with EMT in the microenvironment can be detected. For example, factors increasing angiogenesis were observed to be elevated in the extracellular environment during the EMT process induced by the pleiotropic Y-box-binding protein 1 (YBX1/YB-1) and may play a role in endothelial cell interactions (Gopal et al., 2015).

A PTM intrinsically causes a protein mass shift that can be captured by MS-based proteomics, making proteomics a well-suited methodology to study protein PTMs. Protein phosphorylation is ubiquitous in cells to regulate cellular functions. In addition to the typical sample preparation procedures of DDA-based bottom-up proteomics, an extra enrichment step for phosphorylated peptides is needed for phosphoproteomic sample preparation (**Figure 3**). Phosphoproteomics has revealed two main characteristics of protein phosphorylation. First, it happens unexpectedly fast, with the maximum phosphorylation reached within seconds. Second, the functional phosphorylation sites are abundant, reflected by the high stoichiometry (Aebersold and Mann, 2016). With the fast development of the phosphoproteomic strategy to render deep coverage of the phosphoproteome, it is estimated that more than 75 % of the proteome (researchers now believe the number could be more than 90 %) could be phosphorylated (Aebersold and Mann, 2016). The efficient identification and quantification of the global phosphoproteome in cells by phosphoproteomics have been utilized to decipher the EMT-dependent drug resistance mechanism in TNBC (Golkowski et al., 2020).

### ***The current problems and future of proteomics***

Bottom-up proteomics has greatly dominated the proteomic field to date, relying on the availability of tractable experimental and computational methods. However, it has inevitable flaws, such as loss of information on PTMs and protein-protein interactions. The interest in top-down proteomics has grown considerably to discover the proteoform systematically (B. Chen et al., 2018). Although most of the state-of-the-art proteomic methods can yield a high number of identified proteins, they fail to characterize proteins with

high sequence coverage. The average amino acid coverage for an identified protein in one sample is around 33 %. In addition, single-nucleotide variants and novel splice junctions, which can be observed in DNA and RNA sequencing, are poorly captured by proteomic sequencing (Alfaro et al., 2021).

Single-cell proteomics has emerged in recent years. On account of the intrinsic characteristics of proteomics, there is no generalized workflow for single-cell proteomics, which is still in its infancy with many obstacles to tackle (Marx, 2019). Nanodroplet processing in one pot for trace samples (nanoPOTS) and single-cell proteomics by mass spectrometry (SCoPE-MS) are two advanced MS-based techniques that enable identifying and quantifying more than 1,000 proteins in a single cell (Alfaro et al., 2021). Other strategies include DNA-facilitated and nanopore-based single-molecule protein sequencing, paving the way for the promising future of single-cell proteomics (Alfaro et al., 2021).

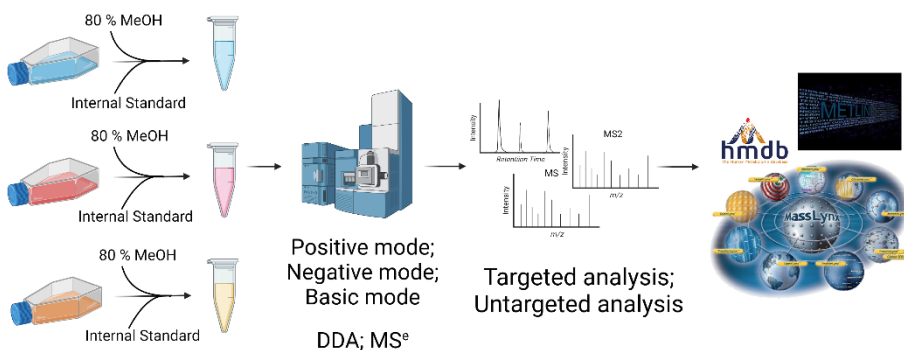
Thanks to the great achievement of proteomic technology and the contribution of the Human Proteome Project (HPP), as of January 2020, more than 90 % of the human proteome had been identified (Overall, 2020). With the flourishing technologies of single-cell proteomics, spatial proteomics, proteome dynamics (protein-protein interactions), and multi-omics, proteomics will help scientists understand the human proteome even better.

### **1.5.2 Metabolomics: A missing puzzle in multi-omics**

The metabolome is the entire collection of small chemicals involved in the cellular metabolic network. These chemical entities or metabolites are not only biomarkers for downstream effects of genes and proteins but also regulators of biological processes. Small metabolites can express their regulatory functions by chemical modification and metabolite-macromolecule interactions. Every popular modification of macromolecules (i.e., DNA, RNA, and protein) requires the participation of small molecules to covalently bind to it, including acetyl-CoA for acetylation, palmitoyl-CoA for palmitoylation, succinyl-CoA for succinylation, SAM for methylation, UDP-Glc and UDP-GlcNAc for glycosylation, and more. A list of metabolites, such as lysine and glutamine, can act as riboswitches to turn on or off the translational process and activate or inhibit specific proteins (Rinschen et al., 2019). Abnormal levels of specific metabolites are closely linked to carcinogenesis, and this group of metabolites is termed oncometabolite. 2-HG, succinate, and fumarate are typical oncometabolites that have been found accumulated in cancer (Rinschen et al., 2019). Oncometabolites do not simply serve as

biomarkers to indicate the dysregulation of certain enzymes; rather, they have regulatory functions through a variety of epigenetic modifications and PTMs. One of the examples of metabolites with bioactivity is fumarate, which reacts with the thiol groups of the cysteine residues in KEAP1, a key regulator of the transcription factor NRF2, to inhibit the normal function of KEAP1 (Adam et al., 2011). NRF2 is a critical regulator for glutathione synthesis and affects cellular oxidative stress, which is associated with cancer malignancy (Pillai et al., 2022).

The revolutionized concept that the metabolome represents not only the readout of the genome and proteome but also intimately controls gene and protein behaviors has urged scientists to integrate metabolomics into the multi-omics strategy. Metabolomics is the methodology to identify systematically bioactive metabolites with either discovery-oriented untargeted or broad-scale targeted profiling analyses. It starts with the identification of metabolite features with specific mass-to-charge ratios ( $m/z$ ) on an LC-MS platform, followed by data mining including statistical analysis and metabolite identification based on spectral libraries and public metabolite databases (**Figure 4**). Ogrodzinski and colleagues recently combined genomics and metabolomics and identified that MMTV-MYC-driven tumors with EMT properties preferentially used the nucleotide salvage pathway, whereas tumors with the papillary subtypes favored *de novo* nucleotide biosynthesis, demonstrating that different types of breast cancer have distinct metabolic phenotypes, a factor that could be exploited to develop subtype-specific therapeutics (Ogrodzinski et al., 2021).



**Figure 4.** General workflow of a typical metabolomics study.

A typical metabolomic study workflow comprises metabolite extraction, LC-MS/MS analysis, metabolite identification, and metabolite quantification. An internal standard metabolite mix that includes a list of well-detected metabolites with heavy isotope labels is added into the samples for quality control. The extracted metabolite samples are then analyzed in the LC-MS/MS system with pre-defined modes and acquisition

methods. Identified peak features are annotated with the human metabolome database (hmdb) and Metlin. The MassLynx software from Waters is used for metabolite quantification. DDA: data-dependent acquisition; MS<sup>e</sup>: data-independent acquisition; LC: liquid chromatography; MS: mass spectrometry.

## 1.6 Our previous work on the metabolism of EMT

Metabolomic data can be integrated with transcriptomic and proteomic data to investigate systematically biological activities, which is exemplified by genome-scale metabolic network reconstructions (GEMs). In GEMs, the gene-protein-metabolic reaction relationships for all the metabolic genes in a study entity (e.g., *Escherichia coli* or mammalian cells) are computationally constructed to simulate and predict metabolic fluxes in different study conditions. To date, the *in silico* reconstruction of metabolic phenotypes via GEMs is best studied in microbes (Fang et al., 2020). Isotope-based tracing (flux analysis)—for example, <sup>13</sup>C- or <sup>15</sup>N-assisted tracing, can be used to track the fate of a specific metabolite and the activity of the associated enzymes in a time-dependent manner to determine the activeness of a biochemistry pathway.

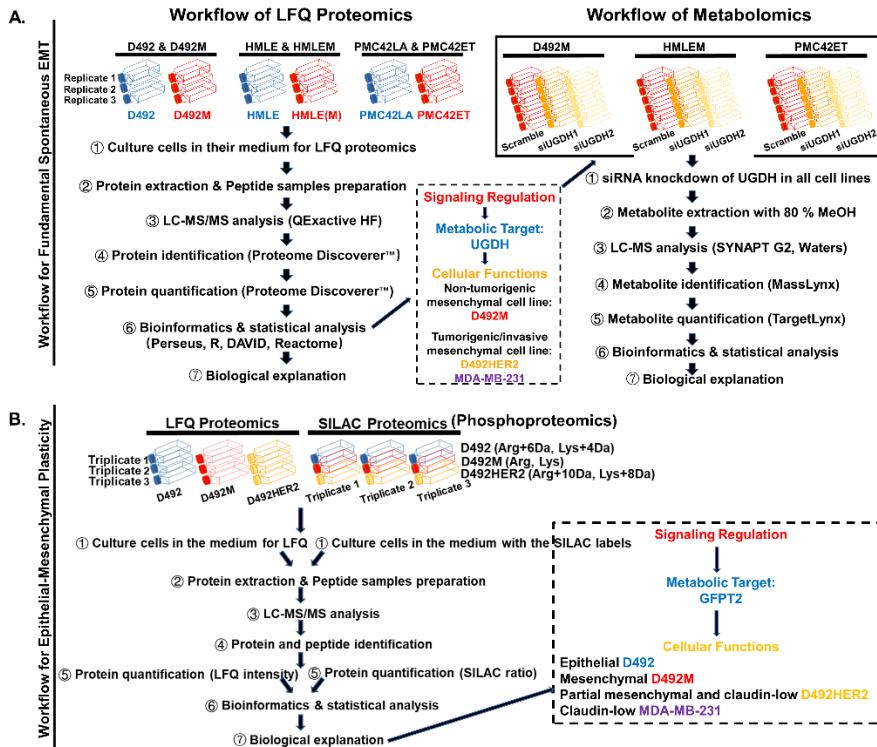
In our lab, we have employed various approaches including <sup>13</sup>C tracing with <sup>13</sup>C-labeled glucose and glutamine, GEMs, targeted and untargeted metabolomic identification and quantification, and lipid profiling to explore the metabolic changes post-EMT. We have observed that epithelial and mesenchymal cells have distinct lipidomic profiles in the D492 and HMLE EMT models. PtdCho and triacylglycerol (TAG) are increased after EMT, while PtdCho- and phosphatidylethanolamine (PtdE) plasmalogens and diacylglycerols (DAG) are decreased in mesenchymal cells. The fatty acids are, on average, shorter and more unsaturated in mesenchymal cells, a phenomenon that may be associated with increased membrane fluidity (Eiriksson et al., 2018). GEMs are constraint-based computational modeling of the cellular metabolic phenotypes. The epithelial and mesenchymal models of D492 were built based on the RNA-seq and microarray analyses of the cellular transcriptome and the quantification of the uptake and secretion of 43 metabolites. Interrogation of these *in silico* models revealed that D492 epithelial cells possess more glutathione, consume more glucose, and secrete more lactate, while the amino acid anaplerosis and fatty acid oxidation fuel the mesenchymal D492M cells (Halldorsson et al., 2017). Recently, with the help of <sup>13</sup>C labeling, we have found that mesenchymal cells can increase the flux via pyruvate carboxylase to replenish the TCA cycle and escalate the flux to citrate formation via reductive carboxylation, while D492 cells mainly oxidize glutamine in the TCA cycle. In addition,

glucose contributes more to fatty acid synthesis than glutamine in D492 cells, while the opposite occurs in D492M cells (Karvelsson et al., 2021).

## 1.7 Project summary

In my Ph.D. project, I focused on the metabolic definition of EMT in breast cancer by using breast EMT cell lines with proteomic approaches supported by metabolomics and functional analyses (**Figure 5**). This project was intended to define the metabolic phenotypes in the process of fundamental spontaneous EMT (**Figure 5A**) and that of EMP or partial EMT (**Figure 5B**). The D492 (D492-D492M), HMLE (HMLE-HMLEM), and PMC42 (PMC42LA-PMC42ET) breast EMT cell models are suitable candidates for the metabolic definition of the essential EMT process on account of their non-tumorigenic properties and the spontaneously induced mesenchymal traits. The characteristics of these cell lines enabled focusing on the fundamental EMT process without the influences of any specific tumorigenic/carcinogenic dominators nor the influences of possessing one or several dominating EMT-related pathways to ensure the plasticity and intrinsic traits of EMT. The isogenic D492 cell lines (D492, D492M, D492HER2) residing at different positions on the EMT spectrum were selected to study the metabolism of EMP. D492 and D492M cells present typical epithelial and mesenchymal phenotypes, respectively, with no tumorigenic abilities, while D492HER2 cells have gone through EMT and possess partial mesenchymal features and the potential to invade.

In this report, I first generally characterized the breast EMT cell lines used in this project to study EMT and EMP, followed by extensive proteomic analyses, which revealed a list of EMT and EMP markers and metabolic enzymes significantly changed post different types of EMT. Next, I identified and further studied the enzymes UGDH and GFPT2, which were representative of the metabolic changes in the rudimentary EMT and partial EMP processes. In recent years, both UGDH and GFPT2 have drawn the attention of oncologists and have been reported to be associated with EMT and survival of patients with cancer. I confirmed that both UGDH and GFPT2 are tumor promoters dysregulated in mesenchymal cells and further explored the upstream signaling regulations and downstream cellular functions of these two enzymes beyond their catalytical activities. Finally, I analyzed the SILAC phosphoproteomic data to define the altered phosphorylation phenotypes among the mesenchymal cell lines D492M and D492HER2 and the epithelial cell line D492 (unpublished data).



**Figure 5.** An overview of the study workflows in this project.

The study workflows illustrate the breast cell lines and main methodologies used in this project, and this figure contains two sections: (1) investigation of the fundamental spontaneous EMT (**A**); (2) investigation of the EMP or partial EMT (**B**). (**A**) In the first section, the proteomes of D492 (D492-D492M), HMLE (HMLE-HMLEM), and PMC42 (PMC42LA-PMC42ET) EMT cell models (or epithelial-mesenchymal cell line pairs) are detected by an LFQ proteomic approach. The topmost dysregulated metabolic enzyme detected in all three EMT models is UGDH. Of which the upstream signaling regulation and downstream functions are investigated. A metabolomic study is conducted to study the metabolic effects of UGDH by siRNA-mediated knockdown of UGDH in three mesenchymal cell lines D492M, HMLEM, and PMC42ET. (**B**) In the second section, the proteomes of the D492, D492M, and D492HER2 cell lines are analyzed by single-shot LFQ supported by a SILAC proteomics approach. GFPT2 is one of the top dysregulated enzymes in the partial mesenchymal D492HER2 cell line compared with the other two cell lines: D492 and D492M. Based on the literature, GFPT2 is part of the mesenchymal metabolic signature (Shaul et al., 2014). To study further, the cells are treated with siRNAs targeting GFPT2 to investigate the upstream signaling regulation and downstream functions of GFPT2. MDA-MB-231 is employed to study the function of UGDH and GFPT2 along with the other D492 cell lines. The phosphoproteomes of D492, D492M, and D492HER2 are analyzed by SILAC, which succeeds in a broad coverage of the phosphorylated proteins (unpublished data). Adapted from (Q. Wang et al., 2021; Q. Wang et al., 2021). Adapted with permission.





## **2 Aims**

The main objective of this Ph.D. project was to characterize the EMT metabolic phenotypes in the human breast gland by implementing proteomic approaches.

The non-binary characteristic of EMT triggered the second objective of this project to define the breast EMP on the proteomic level with a focus on metabolism.

This project further aimed to interrogate the interaction between signaling regulation and metabolism in EMT and EMP via a SILAC phosphoproteomics approach.

The final goal was to identify and characterize specific metabolic targets in EMT and EMP to facilitate breast cancer therapeutic development. The specific objectives in the final goal included confirming the roles of the metabolic enzymes identified in promoting cancer malignancy (including cell proliferation, migration, invasion, and EMT), investigating the canonical and non-canonical functions of the enzymes, and understanding their upstream regulations.



### 3 Materials and methods

All cell lines used in this project were kindly provided by the Stem Cell Research Unit at the Faculty of Medicine, University of Iceland. The D492, D492M, D492HER2, and D492DEE cell lines were established by the Stem Cell Research Unit. The HMLE cell lines (HMLE and HMLEM) were originally from the publication by Prof. Robert A. Weinberg's lab. The PMC42 cell lines (PMC42LA and PMC42ET) were provided originally by the University of Queensland.

#### 3.1 Cell culture

The in-house generated D492, D492M, D492HER2, and D492DEE cell lines were cultured in serum-free H14 medium, which contained the base medium Dulbecco's modified Eagle's medium-F12 (DMEM/F12 without glutamine, Thermo, 21331020) supplemented with 250 ng/ml insulin (Merck, I6634), 10 µg/ml transferrin (Merck, T2252), 10 ng/ml EGF (PeproTech, AF-100-15), 2.6 ng/ml Na-selenite (BD Biosciences, 354201),  $10^{-10}$  M estradiol (Sigma, E2758),  $1.4 \times 10^{-6}$  M hydrocortisone (Sigma, H0888), 0.15 IU prolactin (PeproTech, 100-07), 100 IU penicillin & 0.1 mg/ml streptomycin (Gibco™, 15140122), and 2 mM glutamine (Thermo, 25030024). The MDA-MB-231 cell line was cultured in RPMI 1640 (with HEPES, L-glutamine, and phenol red, Thermo, 52400025) supplemented with 10 % fetal bovine serum (FBS, Gibco™ 10270106) and 100 IU penicillin & 0.1 mg/ml streptomycin. The HMLE cell lines (HMLE and HMLEM) were cultured in DMEM/F12 supplemented with 10 µg/ml insulin, 10 ng/ml EGF,  $1.4 \times 10^{-6}$  M hydrocortisone, 100 IU penicillin & 0.1 mg/ml streptomycin, and 2 mM glutamine. The PMC42 cell lines (PMC42LA and PMC42ET) were cultured in RPMI 1640 medium supplemented with 10 % FBS and 100 IU penicillin & 0.1 mg/ml streptomycin.

DMEM-F12 was replaced by "DMEM:F-12 for SILAC" (Thermo, 88370) with "light-", "medium-", or "heavy-" labeled arginine or lysine (Cambridge Isotope Laboratories) in the SILAC-labeling proteomic experiment. Specifically, the light label (L-arginine, L-lysine), medium label (L-arginine- $^{13}\text{C}_6$  hydrochloride (Arg+6 Da), L-lysine-4,4,5,5-d4 hydrochloride (Lys+4 Da)), and heavy label (L-arginine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$  hydrochloride (Arg+10 Da), L-lysine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_2$  hydrochloride (Lys+8 Da)) were applied for the SILAC experiment.

In the  $^{13}\text{C}$ -tracing experiment,  $^{13}\text{C}$ -labeled 1,2-glucose,  $^{13}\text{C}$ -labeled 1-glutamine, and  $^{13}\text{C}$ -labeled 5-glutamine (Cambridge Isotope Laboratories) replaced the non-labeled glucose or glutamine in the medium, and the base medium used was “DMEM, no glucose, no glutamine, no phenol red” (Thermo, A1443001). In the siRNA-mediated transient knockdown experiment, penicillin and streptomycin were excluded from the cell culture according to the vendor’s instruction. The serum-free H14 medium was supplemented with 10 % FBS in the lower chamber in the invasion assay. All the cell lines were cultured at 37 °C with 5 %  $\text{CO}_2$  for routine maintenance, and cells were routinely checked for mycoplasma contamination.

## **3.2 LFQ proteomics**

### **3.2.1 Protein sample preparation**

D492, D492M, D492HER2, and D492DEE were cultured in T75 flasks in triplicates (three T75 flasks for each cell line) with a seeding density of 600,000 cells/flask. 72 hours after cell seeding, proteins were harvested at around 90 % confluency. Specifically, cells were first washed twice with ice-cold phosphate-buffered saline (PBS) then lysed by 450  $\mu\text{l}$  lysis buffer containing 4 % sodium dodecyl sulfate (SDS, MP Biomedicals™) in 100mM Tris (Sigma). Flasks were then kept on ice for 10 min. Next, the cell lysates were transferred to 1.5 ml Eppendorf tubes and spun at 20,718 rcf for 20 minutes at 4 °C after five freeze (-80 °C)/thaw (room temperature, RT) cycles. After centrifugation, the supernatant was collected and aliquoted in new tubes and stored at -80 °C. Later, the protein concentration was quantified with BCA protein quantification assay (Pierce™). The HMLE and PMC42 cell lines followed the same procedures for the protein sample preparation.

### **3.2.2 Peptide sample preparation**

In the first LFQ experiment containing the D492 (D492 and D492M), HMLE (HMLE and HMLEM), and PMC42 (PMC42LA and PMC42ET) cell lines, 12-15  $\mu\text{g}$  of the total protein was precipitated by chloroform/methanol, reconstituted in 50 mM ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ), reduced with 1 M dithiothreitol (DTT) and 200 mM iodoacetamide (IAA), and digested with 1.5  $\mu\text{g}$  trypsin. Peptides were desalted using C18 STAGETIP after tryptic digestion as described in the literature (Rappsilber et al., 2003), after which peptides were vacuum-dried in a SpeedVac and resuspended in 0.1 % formic acid (FA) for injection.

The peptide preparation procedure for the second LFQ experiment containing D492, D492M, D492HER2, and D492DEE followed the filter-aided sample preparation (FASP) protocol where total proteins equivalent to 300  $\mu\text{g}$  in 150  $\mu\text{l}$  from each sample were reduced with 100 mM DTT, followed by sample processing based on the FASP protocol (Wiśniewski et al., 2009). Proteins were digested twice with trypsin (3  $\mu\text{g}$  x 2) on the filters at 30 °C. The first digestion was done overnight, then for another 6 hours in a final volume of 200  $\mu\text{l}$ . The digested peptides were further desalted with C18 solid-phase extraction cartridges (Empore, Agilent technologies) and resuspended in 50  $\mu\text{l}$  1 % FA. The peptide quantification was measured with pierce quantitative colorimetric peptide assay (product 23275, Thermo Scientific).

### 3.2.3 LC-MS/MS analysis

In the first LFQ experiment, the peptide analysis was carried out on an LC-MS/MS platform that comprised an Easy-nLC 1200 UHPLC system (Thermo Fisher Scientific) interfaced with a QExactive HF orbitrap mass spectrometer (Thermo Fisher Scientific) through a nanospray ESI ion source (Thermo Fisher Scientific). Peptides were injected into a C18 trap column (Acclaim PepMap100, 75  $\mu\text{m}$  i. d. x 2 cm, C18, 3  $\mu\text{m}$ , 100 Å, Thermo Fisher Scientific) and further separated on a C18 analytical column (Acclaim PepMap100, 75  $\mu\text{m}$  i. d. x 50 cm, C18, 2  $\mu\text{m}$ , 100 Å, Thermo Fisher Scientific). Peptides were separated with a multistep gradient running method with buffer A (0.1 % FA) and buffer B (80 % acetonitrile (ACN,  $\text{CH}_3\text{CN}$ ), 0.1% FA). The gradient started from 2%-10% buffer B in 10 min followed by 10%-50% buffer B in 130 min, and 50%-100% buffer B in 20 min with the final step 20 min with 100 % buffer B. Before the next injection, the HPLC was re-equilibrated with 2 % buffer B. The flow rate was set to 250 nl/min. The eluted peptides were analyzed on QExactive HF mass spectrometer operating in positive ion- and DDA mode with the following parameters: electrospray voltage, 1.9 kV; HCD fragmentation with normalized collision energy, 29; automatic gain control (AGC) target value of  $3 \times 10^6$  for orbitrap MS and that of  $1 \times 10^5$  for MS/MS scans. Each MS scan ( $m/z$  350–1650) was acquired at a resolution of 120,000 FWHM, followed by 15 MS/MS scans triggered for AGC targets above  $2 \times 10^3$ , at a maximum ion injection time of 100 ms for MS and 100 ms for MS/MS scans.

In the second LFQ experiment, the digested peptides by trypsin were first separated using an Ultimate 3000 RSLC nanoflow LC system (Thermo Fisher Scientific). Approximately 130 ng of protein was loaded onto an Acclaim

PepMap100 nanoViper C18 trap column (100  $\mu\text{m}$  inner-diameter, 2 cm; Thermo Fisher Scientific) with a constant flow of 5  $\mu\text{l}/\text{min}$ . Peptides were eluted onto an EASY-Spray PepMap RSLC nanoViper column after trap enrichment (C18, 2  $\mu\text{m}$ , 100  $\text{\AA}$ , 75  $\mu\text{m}$ , 50 cm; Thermo Fisher Scientific) and separated with a linear gradient of 2%-35% solvent B (80 % ACN with 0.08 % FA) and Solvent A (0.1 % FA) over 124 min with a constant flow of 300  $\text{nl}/\text{min}$  and column temperature of 50  $^{\circ}\text{C}$ . The HPLC system was coupled to a linear ion trap orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos, Thermo Fisher Scientific) via an EASY-Spray ion source (Thermo Fisher Scientific). The spray voltage was routinely set to 1.8 kV, and the temperature of the heated capillary was set to 250  $^{\circ}\text{C}$ . Full-scan MS survey spectra (335 - 1,800  $m/z$ ) in profile mode were acquired in the orbitrap with a resolution of 60,000 after accumulation of 1,000,000 ions. The top fifteen peptide ions with the most intensities from the preview scan in the orbitrap were fragmented by collision-induced dissociation (CID, normalized collision energy, 35 %; activation Q, 0.250; and activation time, 10 ms) in the LTQ after the accumulation of 5,000 ions. Maximal filling times were 1,000 ms for the full scans and 150 ms for the MS/MS scans. Precursor ion charge state screening was enabled, and all unassigned charge states and singly charged species were rejected. To improve the mass accuracy, the lock mass option was enabled for survey scans (Olsen et al., 2005). Data were acquired by the Xcalibur software.

### **3.2.4 Peptide and protein identification and quantification**

In the first LFQ experiment for the D492, HMLE, and PMC42 EMT models, proteins were identified and quantified by MS data processing in Thermo Scientific™ Proteome Discoverer™ (PD, version 2.3, Thermo). Preview version 2.3.5 from Protein Metrics Incorporate (Kil et al., 2011) was used to inspect the raw files and determine optimal search criteria. A set of search parameters were set up, namely, (1) enzyme specified as trypsin with maximum two missed cleavages allowed; (2) acetylation of protein *N*-terminal including loss-of-Methionine; (3) oxidation of methionine; (4) deamidation of asparagine/glutamine as dynamic PTM; (5) carbamidomethylation of cysteine as static; (6) precursor mass-tolerance of 10 PPM while fragment mass-tolerance of 0.02 dalton. PD's node (Spectrum Files RC) was set up to query the raw files against the human proteome downloaded from UniProt (homo sapiens, file name UP000005640, date October 2018) with the static modification to recalibrate and detect features with the Minora node. The internal contaminants database was also queried along with the human

proteome using the Sequest search engine available in PD (Eng et al., 1994). For both protein and peptide identifications/peptide-spectra-matches (PSM), the false-discovery-rate (FDR) was set to 1 % for downstream analysis of these PSM; only unique peptides with high confidence were used for the final protein group identification. Peak abundances were extracted by integrating the area under the peak curve. The abundance for each protein group was normalized by the total abundance of all identified peptides at FDR < 1 %. Summed up median values for all unique peptide ion abundances mapped to respective protein using LFQ scaled on all average with Precursor Ion Quantifier node for PD were used (Horn et al., 2016).

The raw data files (D492, D492M, D492HER2, and D492DEE four cell lines in triplicates) obtained for each experiment from mass spectrometry were collated into a single quantitated dataset using MaxQuant (version 1.5.2.8) (Cox and Mann, 2008) and the search engine Andromeda (Cox et al., 2011). Enzyme specificity was set to trypsin, allowing for cleavage *N*-terminal to proline residues and between aspartic acid and proline residues. The other parameters were: (1) variable modifications - methionine oxidation, protein *N*-acetylation, gln → pyro-glu, Phospho (STY), deamidation (NQ); (2) fixed modifications, cysteine carbamidomethylation; (3) database for searching: Uniprot-human-up5640 (160516); (4) LFQ: min ratio count, 2 (5) MS/MS tolerance: FTMS - 10 ppm , ITMS - 0.6 Da; (6) maximum peptide length, 6; (7) maximum missed cleavages, 2; (8) maximum labeled amino acids, 3; (9) FDR, 1 %. LFQ intensities were reported individually for each sample and were given as a relative protein quantitation across all samples. LFQ intensities were represented by a normalized intensity profile generated by algorithms described by Cox (Cox et al., 2014). They form a matrix with the number of samples and the number of protein groups as dimensions. The same setups were applied for the iBAQ quantification in the LFQ experiment.

In both experiments, the valid protein identification was defined as “at least two out of three replicates in one cell line must have at least one identified peptide”. Furthermore, the valid protein quantification was defined as below: “at least two out of three replicates in at least one cell line must have detectable intensity”.

### 3.3 SILAC (phospho)proteomics

#### 3.3.1 Protein sample preparation

For the SILAC proteomics and phosphoproteomics analysis, the cell lines D492M, D492, and D492HER2 were fully incorporated with “light”, “medium”, and “heavy” stable isotopes of arginine and lysine, respectively, before protein collection. Cells were first cultured in T25 flasks with respective isotopes of arginine and lysine to get fully labeled cell populations for D492, D492M, and D492HER2. Specifically, the D492 and D492M cells were cultured in the “medium” and “light” labeled medium respectively for six culture passages to ensure that all the cells had reached to the fully labeled status, while the D492HER2 cells were cultured in the “heavy” labeled medium for five passages. Cells were propagated in T75 flasks (Santa cruz), then cultured in T182 flasks (Santa cruz) in triplicates to harvest enough proteins, and the seeding density was 1,500,000 cells/flask. The cell number for seeding was calculated to be consistent with the first LFQ proteomic experiment. Same protein extraction procedures were carried out for the SILAC proteomic and phosphoproteomic experiments as described in the protein preparation section for the LFQ experiments. The only difference was that the lysis buffer was supplemented with one tablet of PhosSTOP phosphatase inhibitors (Roche) and one tablet of cOmplete mini EDTA-free protease inhibitors (Roche) for protein extraction in the SILAC experiments.

#### 3.3.2 Peptide sample preparation

##### ***Protein digestion with FASP***

The extracted proteins were solubilized in 150  $\mu$ l of the protein lysis buffer, which contained Tris HCl (100 mM, pH 7.6) with 4 % SDS and 100 mM DTT. First, protein extracts were heated at 95  $^{\circ}$ C, and to reduce the viscosity of the lysates, DNA was shredded by sonication. Then, samples were centrifuged and processed following the FASP protocol with some modifications (Wiśniewski et al., 2009). Proteins were alkylated in 100  $\mu$ l IAA at a final concentration of 50 mM for 15 min after lysates were passed through the filters (Nanosep, 10k, PALL Life Sciences) which were washed four times with 200  $\mu$ l 8 M urea in Tris-HCl (100 mM, pH 8) and twice with 200  $\mu$ l 40 mM ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ). Proteins were then digested twice with trypsin (3.3  $\mu$ g x 2) on the filters at 30  $^{\circ}$ C. The first digestion was done overnight followed by another 6 hours in 200  $\mu$ l ammonium bicarbonate at 40 mM. The resulting tryptic peptides were eventually desalted with C18 solid-phase extraction cartridges (Empore, Agilent technologies).



### ***High pH reverse-phase peptide fractionation***

Samples (4 mg) were dissolved in 200  $\mu$ L of 10 mM ammonium formate ( $\text{NH}_4\text{HCO}_2$ ) buffer (pH 9.5), and peptides were fractionated using high pH reverse-phase (RP) chromatography. A C18 column from Waters (XBridge peptide BEH, 130  $\text{\AA}$ , 3.5  $\mu$ m, 4.6 x 150 mm, Ireland) with a guard column (XBridge, C18, 3.5  $\mu$ m, 4.6 x 20 mm, Waters) was used on an Ultimate 3000 HPLC (Thermo-Scientific). Buffer A used for fractionation consisted of 10 mM ammonium formate in milliQ water, and buffer B used for fractionation consisted of 10 mM ammonium formate in 90 % ACN. Both buffers were adjusted with ammonia to pH 9.5. Fractions were collected at 1 min intervals using a WPS-3000FC autosampler (Thermo-Scientific). Column and guard column were equilibrated with 2 % buffer B for 20 min at a constant flow rate of 0.75 ml/min and a constant temperature of 21  $^{\circ}$ C. Samples equivalent to 185  $\mu$ l were loaded onto the column at 0.75 ml/min, and the separation gradient started from 2 % buffer B to 5 % buffer B in 6 min, then from 5 % buffer B to 60 % buffer B within 55 min. The column was washed at 100 % buffer B for 7 min, then equilibrated at 2 % buffer B for 20 min, as mentioned above. The collection of sample fractions started 1 min after injection and stopped after 80 min (80 fractions in total, 750  $\mu$ l per fraction). Each peptide fraction was acidified immediately after elution from the column by adding 20 to 30  $\mu$ l 10 % FA to each tube in the autosampler. The total number of fractions concatenated was set to 10. 96 % of material from each fraction was used for the enrichment of phosphorylated peptides, and 4 % was used for total proteome analysis. The fraction content from each set was dried prior to further analysis in the phosphorylated peptide enrichment step.

### ***Phosphorylated peptide enrichment***

The peptides with phosphorylation modification were enriched using MagReSyn-TiIMAC beads (Resyn Biosciences) and Magnetic Rack (DynaMag-2, Life Technologies). The ratio of tryptic peptides to TiIMAC beads were set at ratio 1:5. Beads were first washed using Magnetic Rack with 80  $\mu$ l 1 % ammonium hydroxide ( $\text{NH}_4\text{OH}$ ) or ammonia, followed with 200  $\mu$ l ACN. TiIMAC beads were equilibrated for 2 min with gentle mixing in 200  $\mu$ l loading buffer consisting of 1 M glycolytic acid, 80 % ACN, and 5 % trifluoroacetic acid (TFA). Dried samples were resuspended in 100  $\mu$ l loading buffer and added to TiIMAC beads, and the mixture was incubated with gentle mixing for 20 min at RT. Samples were then washed with 200  $\mu$ l loading buffer for 2 min successively, followed by three times with 200  $\mu$ l of 1 % TFA in 80 % ACN, and in the end with 200  $\mu$ l of 0.2 % TFA in 10 % ACN. The phosphorylated peptides were eluted from beads using 80  $\mu$ l of 1 %

ammonia three times, with pH immediately lowered to 2 using 10 % FA by gentle mixing. Eluted phosphorylated peptides were pooled together then dried in SpeedVac at RT, finally stored at -80 °C before the LC-MS/MS analysis.

### **3.3.3 LC-MS/MS analysis**

Peptide analysis for the total and phosphorylated proteome was performed on a Velos-Pro orbitrap (Thermo Scientific) mass spectrometer coupled with a Dionex Ultimate 3000 RS (Thermo Scientific). The LC buffers used were the following: buffer A (2 % ACN and 0.1 % FA in milliQ water (v/v)) and buffer B (80 % ACN and 0.08 % FA in milliQ water (v/v)). All fractions from both total and phosphorylated proteome were reconstituted in 50 µl of 1 % FA. An aliquot (10 µL of total proteome while 15 µL of the phosphorylated proteome) of each fraction were loaded onto a trap column (100 µm × 2 cm, PepMap nanoViper C18 column, 5 µm, 100 Å, Thermo Scientific) which had equilibrated in buffer A for 19 min at 10 µL/min. The trap column was washed at the same flow rate for 6 min then switched in-line with a resolving C18 column (75 µm × 50 cm, PepMap RSLC C18 column, 2 µm, 100 Å, Thermo Scientific) at a constant temperature of 50 °C. Peptides were eluted at a constant flow rate of 300 nl/min with a linear gradient from 5 % buffer B to 35 % buffer B within 124 min from the column, which was then washed for 20 min at 98 % buffer B and re-equilibrated for 19 min in 5 % buffer B. LTQ-Orbitrap Velos Pro was operated in DDA positive ionization mode. The source voltage was set to 2.6 Kv, and the capillary temperature was 250 °C.

A scan cycle comprised MS1 scan (range from 335 m/z to 1,800 m/z) in the velos-pro orbitrap followed by 15 sequential dependent MS2 scans (the threshold value was set at 5,000, and the minimum injection time was set at 200 ms) in LTQ with CID. The resolution of the Orbitrap Velos was set at 60,000 after the accumulation of 1,000,000 ions. Precursor ion charge state screening was enabled, with all unassigned charge states and singly charged species rejected. Multistage activation for neutral loss ions was activated only for the analysis of phosphorylated peptides. To improve mass accuracy, the lock mass option was enabled for survey scans. The mass spectrometer was calibrated on the first day that the runs were performed to ensure mass accuracy.

### **3.3.4 Peptide and protein identification and quantification**

The raw data files for D492, D492M, and D492HER2 obtained from the mass spectrometer for each experiment were collated into a single quantitated

dataset using MaxQuant (version 1.5.2.8) (Cox and Mann, 2008) and the search engine Andromeda (Cox et al., 2011). Enzyme specificity was set to trypsin, allowing for cleavage *N*-terminal to proline residues and between aspartic acid and proline residues. The other parameters applied in this study were listed, namely, (1) variable modifications: methionine oxidation, protein *N*-acetylation, gln → pyro-glu, Phospho (STY); (2) fixed modifications: cysteine carbamidomethylation; (3) database for protein searching: Uniprot-human\_dec2017 (171216); (4) heavy label: R10K8, medium label: R6K4; (5) MS/MS tolerance: FTMS - 10 ppm, ITMS - 0.6 Da; (6) maximum peptide length, 6; (7) maximum missed cleavages, 2; (8) maximum labeled amino acids, 3; (9) FDR, 1 %. The peak area of labeled arginine/lysine divided by the peak area of non-labeled arginine/lysine for each single-scan mass spectrum was calculated for each arginine- and/or lysine-containing peptide as the peptide ratios. Peptide ratios for all arginine- and lysine-containing peptides sequenced for each protein were averaged. Data were normalized using 1/median ratio value for each identified protein group per labeled sample. The phosphorylated peptides were normalized using the non-phosphorylated protein 1/median values to correct for mixing errors and compared against the individual non-phosphorylated protein ratio itself to correct for protein regulation interactions.

The iBAQ quantification analysis for the SILAC experiment followed similar setups with several different parameters: (1) variable modifications: methionine oxidation, protein *N*-acetylation, Phospho (STY), deamidation (NQ); (2) database for searching: Homo\_sapiens.GRCh38.pep.all (108481); (3) MS/MS tolerance: FTMS - 20 ppm, ITMS - 0.5 Da.

Valid SILAC quantification for each protein was defined as when two out of three replicates for each sample ratio were generated with valid SILAC ratios. Valid phosphoproteomic quantification for each phosphorylation site was filtered by localization probability > 0.75 in all three replicates for each sample ratio.

### 3.4 Data Availability

The mass spectrometry proteomics data (the first LFQ experiment) have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2018) partner repository with the dataset identifier **PXD024164**.

The mass spectrometry proteomics data (the second LFQ experiment) have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier **PXD025600**.

The mass spectrometry (phospho)proteomics data (SILAC) have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier **PXD025858**.

### 3.5 Transient knockdown with siRNAs and quantitative reverse transcription PCR (RT-qPCR)

Cells were seeded either at 60,000 cells/well in 48-well plates or at 480,000 cells/well in 6-well plates. Prior to cell seeding, plates were coated with respective control siRNA (Silencer™ Select Negative Control, 4390843), GFPT2 target siRNAs (Silencer® Select siGFPT2, s19305 and s19306), GSK3B target siRNAs (Silencer® Select siGSK3B, s6239 and s6241), RELA target siRNAs (Silencer® Select siRELA, s11914 and s11915), UGDH target siRNA (Silencer® Select siUGDH, s409 and s410), and PDGFRB target siRNA (Silencer® Select siPDGFRB, s10240) as well as Lipofectamine™ RNAiMAX Transfection Reagent (Thermo). Cells were transfected at 37 °C and 5 % CO<sub>2</sub> for 48 hours with a final siRNA concentration of 10 nM. The sense and antisense sequences of all the siRNAs were listed in **Table 1**.

**Table 1.** Sequences of the siRNAs used in this project.

No.	Target Gene Symbale	siRNA ID	Sense (5' -> 3')	Antisense (5' -> 3')
1	GFPT2	s19305	GACCGAAUUUCACUACAAAtt	UUUGUAGUGAAAUUCGGUCIt
2	GFPT2	s19306	GAUGAUGUCUGAAGACCGAtt	UCGGUCUUCAGACAUCAUCaa
3	GSK3B	s6239	CUCAAGAACUGUCAAGUAAtt	UUACUUAGACAGUUCUUGAGtg
4	GSK3B	s6241	GCUAGAUACACUGUAACAUAtt	UAUGUUACAGUGAUUCUAGCIt
5	RELA	s11914	CCCUUUACGUCAUCCUGAtt	UCAGGGUAGACGUAAAGGGat
6	RELA	s11915	GGAGUACCCUGAGGCUAUAtt	UAUAGCCUCAGGGUACUCCat
7	UGDH	s409	GGGUAACGGUUGUUGAUGUtt	ACAUCAACAACCCUUACCCtg
8	UGDH	s410	CAACAGCGAUUGGAAUGGAtt	UCCAUCCAAUCGCUGUUGet
9	PDGFRB	s10240	Not available	Not available

Cells were cultured in 48-well plates for 72 hours in the RT-qPCR experiments, followed by total RNA extraction with TRI Reagent™ Solution (Invitrogen™). RNA concentration was determined in NanoDrop One (Thermo). 1,000 ng of RNA were used for cDNA synthesis on the thermal cycler (MJ research, PTC-225, Peltier Thermal Cycler) using High-Capacity cDNA Reverse Transcription Kit (Thermo). Gene expression was measured with SYBR Green (Luna® Universal qPCR Master Mix, NEW ENGLAND BioLabs) on Bio-Rad CFX384 Touch™ Real-Time PCR Detection System (Bio-Rad). Primers were selected either based on literature, from PrimerBank, or designed on Primer3Plus website. Primer sequences for genes studied in this project were listed in **Table 2**.

**Table 2.** Sequences of the primers used in this project.

Genes	Primers	Sequences (5'to3')
GFPT2	Forward	ATCCTTGCTTCGCCAAATGC
	Reverse	TTCAGTATCGTCTTGAGCAC
UGDH	Forward	TTTCTGTGCTGTCCAACCCCTGA
	Reverse	CTCTCTGGCCCTCTGGAGTTTC
CDH1	Forward	ACCACGTACAAGGTCAGGT
	Reverse	GGCATCAGCATCAGTCACTT
CDH2	Forward	CCTGCTTATCCTTGTGCTGA
	Reverse	CCTGGTCTTCTTCTCTCCA
SQOR	Forward	CTTCAGGAAGACAGGGAAGCGA
	Reverse	TAACAGTGAGTTCCGCTCCTG
GSK3B	Forward	GGCAGCAAGGTAACCACAGT
	Reverse	GATGGCAACCGATTCTCCAG
PRCKA	Forward	GAAGAACGTGCACGAGGTGAAG
	Reverse	TCCCAAACCCAGATGAAGTC
PDGFRB	Forward	GCCGAGCAACTTTGATCAACGA
	Reverse	GCAGTTCTTGGAGCCAGAAAC
RELA	Forward	CCAGACCAACAACAACCCT
	Reverse	TCACTGGCAGATCTTGAGC
SNAI1	Forward	ACTATGCCGCGCTCTTTCT
	Reverse	AGTCCTGTGGGCTGATGTG
ACTB	Forward	CTTCCTGGGTGAGTGGAGACTG
	Reverse	GAGGGAATGAGGCAGGACTT

### 3.6 Crystal violet assay

The crystal violet assay was conducted for the normalization of the metabolic measurements in the metabolomic experiment with UGDH knockdown and for the glutathione assay in which cells were counted by this assay. In short, cells were fixed with 100 % cold MeOH and stained with 0.25 % crystal violet (Merck, C.I. 42555). After washing, stained cells were dissolved into 100  $\mu$ l of 10 % acetic acid and measured at 570 nm in the microplate reader (SpectraMax<sup>®</sup> M3, Molecular Devices LLC).

### 3.7 Metabolomics analysis

In the GFPT2 knockdown experiment, D492, D492M, and D492HER2 cells in triplicates were transfected with control siRNA (scramble), target siRNA (siGFPT2), or neither (wide-type) for 48 hours in 6-well plates, then cultured for another 24 hours before metabolite extraction. Intensities of the metabolites were normalized to total protein concentration measured by BCA assay. In the UGDH knockdown experiment, the D492M, HMLEM, and PMC42ET cell lines in sextuplicate were transfected with control siRNA (scramble) and target siRNA (siUGDH) following the same procedures for siGFPT2 with the intensities of the metabolites normalized to the crystal violet signals. In the <sup>13</sup>C-labeling experiments, wide-type cells were cultured in T25 flasks in triplicates, and after cells reached 80 % confluency, the medium was

changed to ones without glucose or glutamine. After culturing cells in the medium deprived of glucose or glutamine for four hours (as Time 0), labeled  $^{13}\text{C}$  1,2-glucose or  $^{13}\text{C}$  1(5)-glutamine was added. Metabolites were extracted at time 0 and after 6 hours.

Before metabolite extraction, cells were first washed with sterile saline solution. Metabolites were extracted with cold 80 % MeOH (MeOH: dH<sub>2</sub>O, 80:20) containing an isotopically labeled internal standard mix (**Table 3**) as instructed in an in-house protocol. Briefly, after adding 80 % MeOH into the cell culture, cells were scraped off, and cell lysates were centrifuged. Next, the supernatant was taken and vacuum-dried. The extracted metabolites were analyzed on the ultra-performance liquid chromatography (UPLC) coupled with ESI qTOF mass spectrometry (SYNAPT G2, Waters) according to published protocols (Rolfsson et al., 2017). Briefly, the metabolites lysates were loaded onto a BEH amide column (Acquity, 1.7  $\mu\text{m}$ , 2.1 x 150 mm, Waters) to achieve chromatographic separation by hydrophilic interaction liquid chromatography (HILIC). Two running conditions (acidic and basic) and two ionization modes (negative and positive) were used for the chromatographic separation, including acidic negative, acidic positive, and basic negative modes. The running buffers for the acidic condition included buffer A (0.1 % FA in ACN (v/v)) and buffer B (0.1 % FA in milliQ water (v/v)), while for the basic condition, the mobile phases were buffer A (100 % ACN) and buffer B (20 mM ammonium acetate (pH 9.4, NaOH)).

The metabolite identification and quantification were conducted in MassLynx software (Waters, version 4.2) from Waters. The targeted metabolites were identified based on an in-house built library, which was generated by identifying expected metabolites with chromatographic retention time, accurate mass, and their ion adducts and fragments. For untargeted metabolites identification, the publicly deposited spectrum from the online human metabolome database (HMDB) (Wishart et al., 2018) and METLIN (C. A. Smith et al., 2005) were used to cross-reference the data collected from this project. The integration of the area under the peak curve for each targeted metabolite was conducted in TargetLynx (Waters, version 4.2).

For the  $^{13}\text{C}$ -labeling tracer experiments, data were analyzed in TargetLynx, and IsoCore was used to correct for the abundances of naturally heavy isotopes (Millard et al., 2012). We normalized the mean enrichment of  $^{13}\text{C}$  in UDP-GlcNAc to the total amount of UDP-GlcNAc and presented it as relative  $^{13}\text{C}$  incorporation.

**Table 3.** The composition of the internal standard mix.

The isotopically labeled metabolites in the internal standard mix for the metabolic experiments were listed. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

No.	Internal Standards	Concentrations ( $\mu\text{g/mL}$ )
1	Adenine ( $^{15}\text{N}_2$ )	50
2	Alanine (d4)	1000
3	AMP ( $^{13}\text{C}_{10}$ , $^{15}\text{N}_5$ )	50
4	Arginine ( $^{13}\text{C}_6$ )	50
5	Carnitine (d9)	20
6	Citric acid ( $^{13}\text{C}_6$ )	50
7	Cysteine ( $^{13}\text{C}_3$ , $^{15}\text{N}$ )	50
8	Glucose ( $^{13}\text{C}_6$ )	2100
9	Glutamic Acid (d5)	30
10	Glutamine ( $^{15}\text{N}_2$ )	50
11	Lysine (d4)	90
12	Malonic acid (d4)	50
13	Octanoic Acid (d15)	150
14	Phenylalanine (d2)	72
15	Phtalic Acid (d4)	50
16	Succinic acid (d4)	50

**Table 4.** The main metabolites identified and quantified in this project were listed.

The first four metabolites were involved in the EMT section in this project (paper I), while the rest of the metabolites were involved in the EMP section (paper II).

No.	Metabolites	Mass	Retention Time	Mode
1	UDP-Glucose	565.0476	8.47	Basic, Negative
2	UDP-Glucuronic acid	579.028	8.58	Basic, Negative
3	Glycerophosphocholine	109.041+242.078	4.79	Acidic, Negative
4	Acetylaspartate	174.03868	2.83	Acidic, Negative
5	UDP-N-acetylglucosamine	606.074	6.58	Acidic, Negative
6	Glutamate	146.0459	4.35	Acidic, Negative
7	Cystathionine	221.0596	5.38	Acidic, Negative
8	Reduced Glutathione	306.077+272.09	4.28	Acidic, Negative
9	N-acetylglucosamine-1-phosphate	300.0484	5.21	Acidic, Negative

### 3.8 Untargeted metabolomics analysis

The R package *XCMS* was used in the untargeted data analysis (Colin A. Smith et al., 2006) for automatic chromatographic peak-picking (the *centWave* algorithm) (Tautenhahn et al., 2008) and retention time alignment

between samples (the *OBI-Warp* algorithm) (Tsugawa et al., 2015). Features that eluted in the first 66 seconds were omitted from further analysis. Feature intensities were normalized by quality-control sample-based robust *LOESS* (locally estimated scatterplot smoothing) signal correction (QC-RLSC) (Dunn et al., 2011), which was implemented by the R package *NormalizeMets* (De Livera et al., 2018). All features with over 25 % relative standard deviation (RSD) in the QC samples were omitted from further analysis for quality assurance. To obtain mean-centered and normally-distributed feature intensity values with equal variance, the generalized logarithmic transformation (*glog*) (Durbin et al., 2002) and autoscaling were applied.

### **3.9 Cell proliferation assay**

Cells were seeded in 96-well plates at 10,000 cells/well in quadruplicates. Both GFPT2 and UGDH knockdowns followed the protocol described in the above section. In the GFPT2 knockdown experiment, 24 hours after cell seeding for D492 and D492M while 48 hours for D492HER2, cells were placed under the microscope (LEICA CTR 6500, bright field, 10x) with 5 % CO<sub>2</sub> at 37 °C for real-time monitoring and multiple data acquisition. This was controlled by software Micro-Manager (version 1.4.22). Three spots were chosen in each well, and photos were taken every 6 hours. Cell growth was monitored for 66 hours for D492 and D492M while 42 hours for D492HER2. Photos were batch-processed with Macro in software ImageJ 1.52p, and cell numbers were normalized to the starting time point for monitoring under the microscope. The same procedures were performed in the UGDH knockdown experiment for the D492M and D492HER2 cell lines.

### **3.10 Scratch migration assay**

In the GFPT2 knockdown experiment, the scratch assay was performed in the IncuCyte ZOOM system (2018A) following the manufacturer's instructions. Cells in triplicates were seeded at 40,000 cells/well in 96-well plates (Essen bioscience, ImageLock, 4379). GFPT2 knockdown followed the procedures described in the above section. Briefly, cells were scratched and put into the IncuCyte after 48 hours of transfection with siRNAs. The IncuCyte ZOOM system took pictures every 2 hours. Two positions in each well were chosen, and cells were monitored for around 72 hours to reach full wound closure. Images were analyzed in the software IncuCyte ZOOM (2018A), and the wound confluence data were exported from the same software.



### 3.11 Transwell invasion assay

The D492M and D492HER2 cells were cultured with siRNA transfection (Scramble, siGFPT2, and siUGDH) for 48 hours in a 6-well plate, and the protocol for the GFPT2 and UGDH knockdown procedure was described above. Cells were then re-seeded into filter units (Falcon® Permeable Support for 24-well Plate with 8.0 µm Transparent PET Membrane, 353097) coated with Matrigel (Corning® Matrigel® Matrix, 356234) at a density of 30,000 cells/well. Firstly, the filter inserts were coated with 100 µl 1:10 diluted Matrigel for 20-30 min at 37 °C. Secondly, 300 µl of cell suspension was added on top of the filter units. Thirdly, 500 µl of H14 medium with 10 % FBS were added to the wells in the 24-well plates below the filters. Finally, cells were incubated at 37 °C and 5 % CO<sub>2</sub> for 48 hours. To normalize the different cell numbers in the filter units, cells were seeded into a 24-well plate along with the filter units and cultured in the same way as cells in the filter units.

To count the invasive cells below the filter units, non-invasive cells on top of the filters were first removed with cotton swabs, followed by fixation with paraformaldehyde (PFA, 3.7 %, Sigma, 252549) and DAPI staining (1:5000, Sigma, D9542). 10 images per filter unit were then taken by the EVOS® FL Auto Imaging System (10x, Thermo), followed by the batch analysis of the images in Macro ImageJ 1.52p.

### 3.12 Western blot

The D492 cells were incubated with the GFPT2 siRNAs as described above in the knockdown experiment. Protein lysates were extracted with RIPA buffer (Pierce™, 89900, Thermo) supplemented with protease & phosphatase inhibitors (Halt™, 1861284, Thermo) and quantified with BCA protein assay. Proteins were separated by the 4%-12% Bis-Tris gels (NuPAGE™, Thermo), transferred to polyvinylidene fluoride (PVDF) membranes (IPFL00010, Immobilon®), and probed with antibodies against O-GlcNAcylation (1:200 dilution; sc-59623; Santa Cruz Biotechnology) and the loading control β-actin (1:2,000; MA5-15739; Thermo). The western blot detection reagents were Clarity Max™ Western ECL substrate (Bio-Rad), and plots were imaged in the Molecular Imager® ChemiDoc™ XRS+ Systems (Bio-Rad). The protein abundance between samples was quantified and compared via densitometry quantification of western blots carried out in the software ImageJ 1.52p.

### **3.13 Hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and reduced glutathione (GSH) treatment and growth factor deprivation**

MDA-MB-231 cells were seeded in 24-well plates at 300,000 cells/well and cultured for 48 hours, followed by treatment with 2 $\mu$ M H<sub>2</sub>O<sub>2</sub> (Honeywell, 18304H) for 2 hours.

MDA-MB-231 cells were seeded in 24-well plates at 200,000 cells/well and cultured for 24 hours, followed by treatment with 50 mg/L reduced glutathione (GSH, Sigma, G4251) for 48 hours. Cells were changed with fresh GSH-containing medium 2 hours before the RNA extraction.

MDA-MB-231 cells were cultured in the H14 medium as described for the D492 cell lines. The cells were then seeded in 24-well plates at 200,000 cells/well and cultured for 24 hours, followed by treatments with medium deprived of insulin or EGF for 48 hours. The fresh medium was changed for the cells 2 hours before the RNA extraction.

In the above experiments, the GFPT2 gene expression was tested by RT-qPCR as described in the above section.

### **3.14 Glutathione assay**

The glutathione levels were measured with the GSH/GSSG-Glo™ Assay from Promega (V6611), covering both reduced (GSH) and oxidized (GSSG) glutathione. Cells in quadruplicates were seeded at 20,000 cells/well in 96-well plates. GFPT2 knockdown and H<sub>2</sub>O<sub>2</sub> treatment followed the protocols described in the above section. The glutathione levels were measured 24 hours after the medium was changed. The luminescence signal was detected in the microplate reader (SpectraMax® M3, Molecular Devices) with white and opaque 96-well plates (BRANDplates®, 781965). To normalize the glutathione level, cells were counted using a crystal violet assay as described in the previous section.

### **3.15 Statistical analysis and bioinformatics**

All experiments performed in the first study on EMT and UGDH (paper I) were at least in triplicates. The metabolomics analysis of the UGDH knockdown in D492M, HMLEM, and PMC42ET was in sextuplicate. The proteomic data were analyzed in Perseus (Version 1.6.14.0, data imputation based on a normal distribution with width as 0.3 and down shift as 1.8, Permutation-based FDR < 0.05) (Tyanova et al., 2016) and R (Version 4.0.0). The figures reported in this study were generated in R. The statistical

significance for all the two-sample comparisons was derived from the two-sided Student's T-test (Welsch,  $p < 0.05$ ). All error bars represented standard deviation (SD). The Gene Ontology (GO) functional annotation was carried out in the DAVID online platform (DAVID Bioinformatics Resources 6.8) with default settings (Huang da et al., 2009; Huang da et al., 2009). The Reactome pathway analysis was conducted with the online Reactome tool (Pathway browser version 3.7; Reactome database release: 75) with default settings (Jassal et al., 2019). Proteins with significance above a certain threshold (permutation-based FDR  $< 0.05$ ) were included in the GO annotation and Reactome pathway analysis. Patient survival analyses were conducted by the online tool kmplot.com (basal breast cancer patients, split patients by auto select best cutoff) (Györfy et al., 2010).

In the second study on EMP and GFPT2 (paper II), we utilized two quantitative methods for proteomics analysis (single-shot LFQ and ten-fraction SILAC) to increase the data validity and the reproducibility of the findings. In the LFQ experiment, four cell lines (D492, D492M, D492HER2, and D492DEE) in three biological replicates (12 samples in total) were analyzed, and the proteins with statistical significance were thoroughly described. In the SILAC experiment, three cell lines (D492, D492M, and D492HER2) in three biological replicates (nine samples in total) were analyzed and described. Due to the limitations with the SILAC methods (maximum labeling capacity is three), the D492DEE cell line was expelled from the SILAC experiment. The epithelial D492 cell line was used as the control for the D492M and D492HER2 mesenchymal cell lines in both LFQ and SILAC. Statistical analysis for all the comparisons between different treatments was carried out in R and Perseus (two-sided one or two sample(s) Student's T-test). All error bars represented SD.

The R packages "ComplexHeatmap", "ggdendro", and "dendextend" (Galili, 2015; Z. Gu et al., 2016) were used to generate the heatmaps and dendrogram in this project. The volcano plots were plotted in R using data that were first analyzed and organized in Perseus (version 1.6.2.3, replace missing values from normal distribution, width = 0.3 and down shift = 1.8, two-sided Student's T-test for LFQ, one sample T-test for SILAC, Permutation-based FDR). GO annotation was carried out in Perseus (version 1.6.12.0, Fisher exact test, Benjamini-Hochberg FDR) (Tyanova et al., 2016). All identified proteins from the SILAC experiment were used as background for the GO annotation. The KEGG pathway enrichment was performed by the R package "pathfindR" (100 iterations; Protein-protein interaction: Biogrid; p-values adjustment: "bonferroni", adjusted p-value threshold: 0.05) (Ulgen et

al., 2019). Reactome metabolic pathway enrichment was conducted with the default parameters on the Reactome website (Version 65, 67, and 72 were used for D492 vs. D492M, D492 vs. D492HER2, and D492M vs. D492HER2, respectively) (Jassal et al., 2019) then plotted in R as treemaps. Proteins involved in the metabolic pathways (enrichment FDR < 0.05) were mapped in the protein interaction networks created in STRING (Version 11.0; k-means clustering, minimum required interaction scores: medium confidence 0.400) (Szklarczyk et al., 2019) and visualized in Cytoscape (version 3.5.1/Version 3.6.1) (Shannon et al., 2003). The R packages “survminer” and “survival” were used for the survival analysis in breast cancer patients, and the top and bottom 20<sup>th</sup> percentile of patients were included in the analysis. The data of patients with breast cancer were downloaded from The Cancer Genome Atlas (TCGA) cBioPortal (Breast Invasive Carcinoma (TCGA, Provisional)) (Hoadley et al., 2018). The EMT markers referred to in this project were downloaded from the online EMT marker database (Min Zhao et al., 2015). The GFPT2 RNA expression data in breast cell lines and breast cancer patients were downloaded from the Cancer Cell Line Encyclopedia (CCLE) database (Ghandi et al., 2019), the Harvard Medical School (HMS) LINCS database (Koleti et al., 2017), and TCGA cBioPortal (Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)) (The Cancer Genome Atlas Network, 2012). The proteins identified and quantified in both LFQ and SILAC were plotted as scatter plots in R (Pearson).

In both the EMT and EMP studies, the phosphoproteomic data were analyzed in the Ingenuity Pathway Analysis (IPA) (QIAGEN, version from 2018) for pathway enrichment and in Perseus for motif enrichment analysis.

All the R codes used for analysis and plotting on UGDH can be found on [https://github.com/QiongW56/UGDH\\_Publication\\_2021](https://github.com/QiongW56/UGDH_Publication_2021). (Paper I)

All the R codes used for analysis and plotting on GFPT2 can be found on [https://github.com/QiongW56/GFPT2\\_Publication\\_2021](https://github.com/QiongW56/GFPT2_Publication_2021). (Paper II)

## **4 Results and discussion**

### **4.1 Proteomic profiling of the EMT and EMP programs**

The notion that EMT is a driver for cancer pathogenesis has propelled the drastic growth of the EMT research field. EMT is not a binary process but rather with a continuum of intermediate states to render cells phenotypic plasticity. We employed three epithelial-mesenchymal breast cell line pairs to study the essential non-tumorigenic spontaneous EMT program and three isogenic breast cell lines with plasticity and tumorigenicity to study the invasive EMP program. The background and phenotypes of all the cell lines were introduced, and the proteomic changes of both the EMT and EMP programs by comparing different cell lines were analyzed, followed by functional annotation of the proteomic changes.

#### **4.1.1 Characterize the D492, HMLE, and PMC42 EMT cell models (paper I)**

The D492 (D492-D492M), HMLE (HMLE-HMLEM), and PMC42 (PMC42LA-PMC42ET) EMT cell models were chosen and characterized to study EMT (**Figure 6**). The D492 and HMLE epithelial cell lines were generated from normal breast reduction mammoplasty with different cell immortalization methods. The D492 epithelial cells were basal-like with stem cell properties, while the D492 mesenchymal cells were derived spontaneously from the epithelial cells after coculture with endothelial cells. HMLE cells isolated from breast tissue contained both luminal and basal epithelial cells, and the mesenchymal counterpart was induced spontaneously by instant trypsinization (1 minute). On the contrary, the PMC42 mesenchymal cells were derived from pleural effusion from the metastatic site in a breast cancer patient, and the epithelial counterpart was produced spontaneously by hormone treatment via MET. All the cell lines held certain levels of myoepithelial or basal features of breast epithelial cells and were not tumorigenic.

Genetic manipulation to induce EMT, such as overexpression of EMT-related growth factors, transcription factors, and microRNAs, may lead to unwanted effects that diminish the plasticity and flexibility of the EMT program (Aharonov et al., 2020; Cano et al., 2000; Krebs et al., 2017; Tam et al., 2013; Y. Yu et al., 2018). The three breast EMT cell models symbolize the natural EMT program in human breast gland development and have been

adopted in various studies and contributed indubitably to the understanding of the underlying mechanism of EMT (Bhatia et al., 2019; Briem et al., 2019; Kröger et al., 2019).

	D492-D492M	HMLE-HMLEM	PMC42LA-PMC42ET
Tissue	Breast reduction mammoplasty	Breast reduction mammoplasty	Breast; derived from metastatic site: pleural effusion
Immortalization	HPV-16 E6/E7 gene transduction	hTERT and SV-40 Large T antigen expression	N/A
Cell type	Basal-like; with stem cell properties	Luminal and basal epithelial cells	Express secretory and myoepithelial markers
Induce EMT	Spontaneous EMT; Co-culture with endothelial cells	Spontaneous EMT; Trypsinization	Spontaneous MET; Treatment with a combination of hormones to induce MET
Tumorigenicity	Non-tumorigenic	Non-tumorigenic	Non-tumorigenic

**Figure 6.** Breast spontaneous EMT cell models.

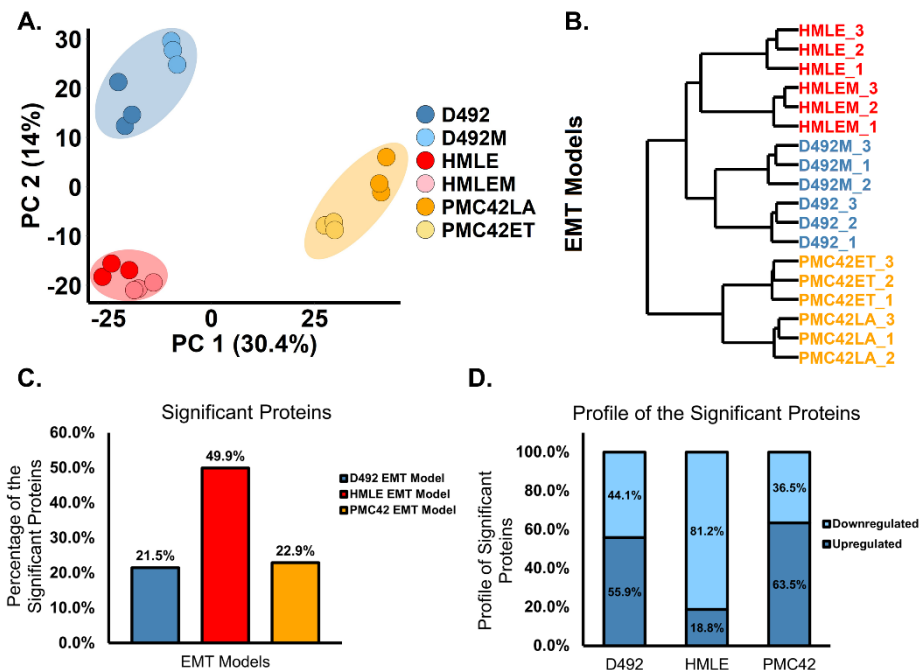
The three breast EMT cell models (epithelial-mesenchymal cell line pairs) used to study the essential spontaneous EMT were introduced and compared in terms of their tissue origin, methods of immortalization, cell type markers, methods of EMT induction, and tumorigenicity *in vivo*. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

Proteomic analysis of the three EMT cell models revealed that cell lines with the same tissue origin were more similar in spite of their epithelial or mesenchymal properties (**Figure 7A**). The proteomes of the PMC42 cell lines were distant from those of D492 and HMLE models (**Figure 7B**). 873 valid proteins in total were identified and quantified in the proteomic analysis. 21.5 % of the valid proteins in the D492 cell model, 49.9 % of those in HMLE, and 22.9 % in PMC42 were significantly altered after EMT (Permutation-based FDR < 0.05) (**Figure 7C**). Within the significantly changed proteins after EMT, 55.9 % in the D492 model, 18.8 % in the HMLE model, while 63.5 % in the PMC42 model were upregulated (**Figure 7D**).

The failure of the epithelial-mesenchymal characteristics being determinant for the cell line clustering suggests that the impacts of the spontaneous switches between the epithelial and mesenchymal states are less profound than the inherited genetic background among the cell lines (**Figure 7A**). Both D492 and HMLE are derived from normal breast tissue,

while PMC42 is isolated from the metastatic site of a breast cancer patient, which may be responsible for the alienation of PMC42 (**Figure 7B**).

The D492 and PMC42 cell models shared similar percentages of the altered proteome after EMT, which was more than two times lower than that of the HMLE proteome (**Figure 7C**). Furthermore, both D492 and PMC42 models had relatively more proteins upregulated after EMT. However, less than 20 % of the changed proteins in the HMLE model were upregulated (**Figure 7D**). It suggests that the HMLE epithelial cell line is more epithelial-like than the other two epithelial cell lines.



**Figure 7.** Comparison of the proteomes among the three breast EMT cell models and during the EMT process.

**(A)** PCA analysis of the proteomes of the three EMT models clustered the cell lines with the same tissue origin together, while the epithelial (or mesenchymal) proteomic phenotypes of these cell lines were not close enough to cluster. **(B)** The proteome of D492 was more similar to that of HMLE than PMC42. **(C)** 21.5 % of the detected proteome in the D492 EMT model were significantly altered after EMT (Permutation-based FDR < 0.05), while 49.9 % in HMLE and 22.9 % in PMC42 were significantly changed. **(D)** 55.9 % of the significantly changed proteome in the D492 EMT model, 18.8 % in HMLE, and 63.5 % in PMC42 were upregulated. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

We further mapped the log<sub>2</sub> fold changes and -log<sub>10</sub> p values of all the detected proteins to filter and identify consistent and significant changes among all the EMT cell models (**Figure 8**). Most of the proteins had failed to yield consistent changes, empathizing the heterogeneity of the EMT program and supporting that EMT is a context-specific biological process.

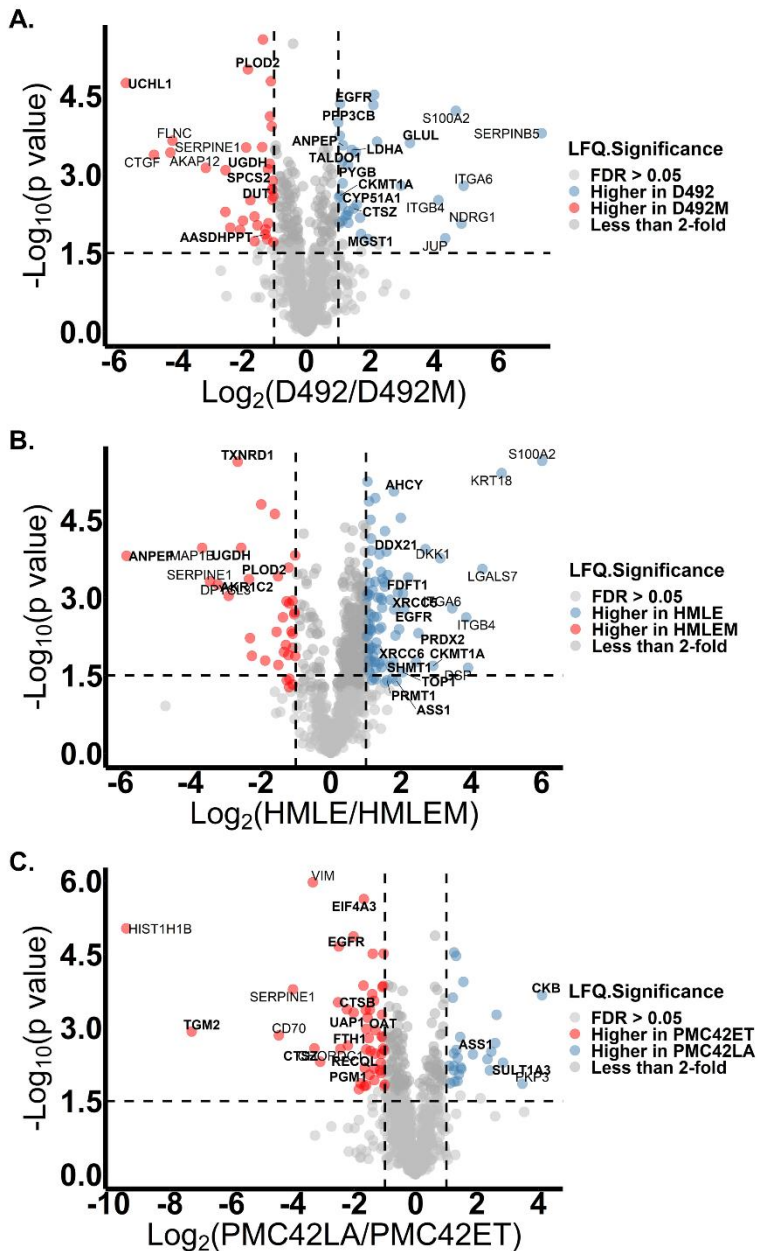
GO annotation of the significantly altered proteins after EMT in the three EMT cell models revealed that changes to the biological process (BP) “cell-cell adhesion” were prevalent among all models (**Figure 9A-C**). The HMLE and PMC42 cell models had changed their biological functions to a similar extent, with the PMC42 model sharing its top seven GO-BP terms with HMLE. The estrangement of D492 from the other cell models in EMT-related functional changes was confirmed by the Reactome pathway enrichment analysis (**Figure 9D-F**). Response to cell stress and alteration in IGF and interleukin-12 signaling were pathways reshaped in the D492 model after EMT, while changes in the protein translational process were mainly involved in the HMLE and PMC42 cell models.

The protein expression profile of the D492 cell lines was similar to that of HMLE compared with PMC42 (**Figure 7B**); however, the altered biological functions post-EMT were discordant between these two EMT cell models (**Figure 9**). In HMLE and PMC42, the epithelial-mesenchymal switch is mainly promoted by the changed translational activities, suggesting that these cells rely on the translational machinery to reshape their proteomic landscapes during the phenotypic switch. The extravagant difference to the altered protein profile between HMLE and PMC42 suggests that the cell lines have customized their needs for protein translation in EMT induction (**Figure 9C-D**). The D492 EMT cell model suffices the transition by altering cell responses to stress and signaling regulations, suggesting that these cells, relying on the cell plasticity imparted by the stem cell-like properties of the D492 epithelial cells, commit to post-translational approaches to fulfill EMT. These findings indicate that certain distinct cell properties, such as stem cell properties, can be dominant for EMT induction, surpassing the proteomic similarity, and cells with discordant proteomes can rely on similar functional mechanisms tailored to their intrinsic characteristics to induce EMT.

The D492, HMLE, and PMC42 EMT cell models are three sets of cell lines with distinct genetic backgrounds and heterogenous EMT processes. D492M, HMLEM, and PMC42ET represent three distinct mesenchymal phenotypes, supporting that the mesenchymal cell state is very diverse and can not be well-defined by certain sets of markers (J. Yang et al., 2020). These

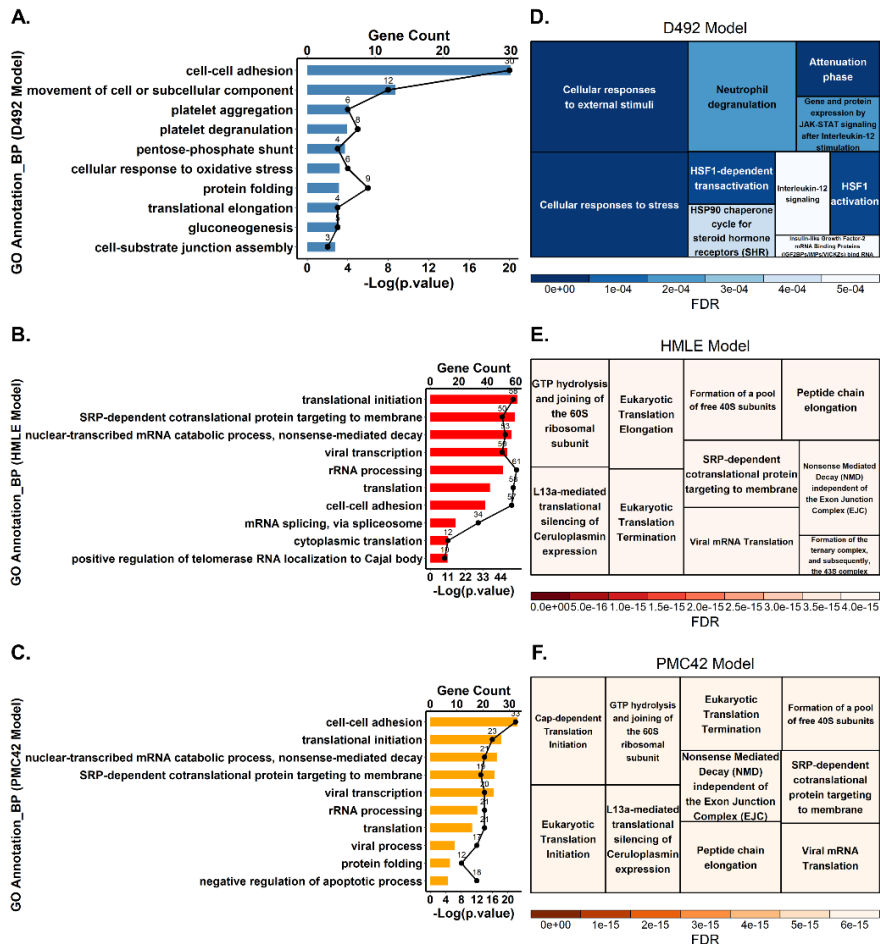


mesenchymal cell types may be subject to further oncogenic insults and eventually cause different levels of oncological severity. These findings support the emerging concept: The intrinsic properties induced by the cell of origin are the foundation and determinant of the subsequent oncogenic events, which can profoundly affect cancer malignancy (Puisieux et al., 2018).



**Figure 8.** Proteomic changes after EMT in the three EMT models.

The log2 ratios (epithelial/mesenchymal) of the detected proteins, along with the  $-\log_{10}(p \text{ value})$ , were plotted for the D492 EMT cell model **(A)**, HMLE EMT cell model **(B)**, and PMC42 EMT cell model **(C)** based on the proteomic analysis. The proteins with FDR (Permutation-based)  $< 0.05$  and fold change  $> 2$  were colored. The p value, which equals to 0.03 ( $-\log_{10}(p \text{ value}) = 1.5$ ), was indicated by the horizontal dash lines, while the fold change at 2-fold was highlighted by the vertical dash lines. The normal label-marked proteins had  $\log_2(\text{fold change}) > 3$ , i.e., UCHL1, CTGF, SERPINE1, FLNC, AKAP12, GLUL, ITGB4, JUP, S100A2, NDRG1, ITGA6, and SERPINB5 in the D492 model **(A)**; ANPEP, MAP1B, SERPINE1, DPYSL3, DKK1, ITGA6, ITGB4, DSP, LGALS7, KRT18, and S100A2 in the HMLE model **(B)**; while VIM, HIST1H1B, SERPINE1, TGM2, CD70, CTSZ, CHORDC1, PKP3, and CKB in the PMC42 model **(C)**. Metabolic proteins with  $\log_2(\text{fold change}) > 1$  for the D492 cell model, 1.5 for the HMLE and PMC42 cell models were bold label-marked, for example, UCHL1, UGDH, GLUL, and LDHA in D492 **(A)**; UGDH, FDFT1, ASS1, and AHCY in HMLE **(B)**; and PGM1, UAP1, OAT, ASS1, and SULT1A3 in PMC42 **(C)**. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



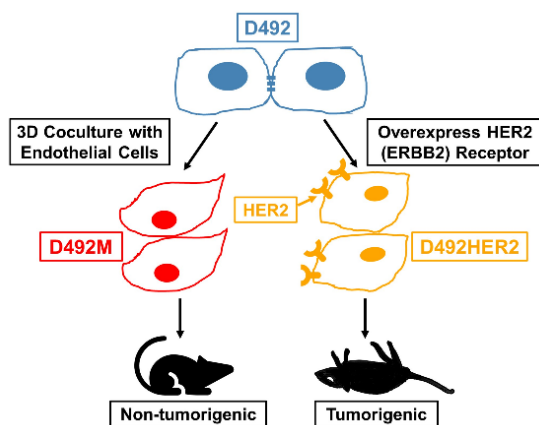
**Figure 9.** Functional changes after EMT in the three EMT models.

**(A-C)** The top 10 of the most dysregulated biological processes (BP) were listed for the D492 cell model **(A)**, the HMLE cell model **(B)**, and the PMC42 cell model **(C)**, respectively. The GO terms were listed according to the  $-\log_{10}(p \text{ value})$  in a descending order. The gene numbers contained in each GO term were plotted as dots/line. The functional annotation of the GO-BP terms was performed on the DAVID platform (DAVID Bioinformatics Resources 6.8). **(D-F)** The Reactome pathway analysis was conducted for the D492 cell model **(D)**, HMLE cell model **(E)**, and PMC42 cell model **(F)** (Pathway browser version 3.7; Reactome database release: 75). Both the GO annotation and Reactome pathway analysis were analyzed with the proteins significantly altered in each EMT cell model (Permutation-based FDR < 0.05). Both the DAVID GO annotation and Reactome pathway analysis were performed with default settings on the platforms. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

#### **4.1.2 Characterize the D492, D492M, and D492HER2 EMP cell model (paper II)**

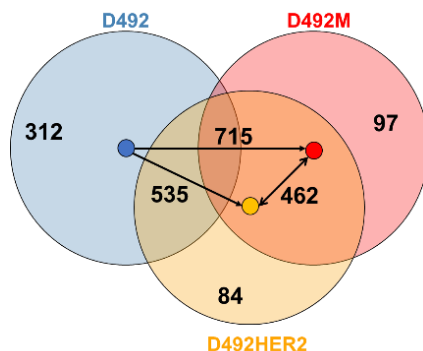
To discriminate proteomic phenotypes that confer cell invasiveness from migration, the cell line trio, epithelial D492, mesenchymal D492M, and partial mesenchymal D492HER2 formed the EMP cell model and was analyzed in the same proteomic experimental setup **(Figure 10)**. D492 is a normal human breast epithelial cell line with basal-like and stem cell properties and is capable of initiating EMT and of differentiating into both luminal epithelial and myoepithelial cells. D492M was generated by spontaneous EMT induction from D492 and is non-tumorigenic with typical mesenchymal phenotypes and with increased migratory ability. D492HER2 cells possess certain mesenchymal traits with more potential for invasiveness, representative of the EMT plasticity.

The proteomic analysis confirmed that D492HER2 is more similar to D492M than D492 on the protein level, with 462 differently expressed proteins identified between D492HER2 and D492M while 535 proteins in D492HER2 different from D492 **(Figure 11)**. The significant difference to proteome between D492M and D492 (715 altered proteins) exceeded that between D492HER2 and D492 (535 altered proteins), confirming that D492HER2 locates in-between the epithelial D492 and mesenchymal D492M, retaining the epithelial-mesenchymal intermediate phenotype. The smaller number of signature proteins (signature proteins: a set of proteins uniquely dysregulated in one cell line compared with the other two cell lines) in D492HER2 (84 proteins) further confirmed that D492HER2 is an intermediate state between D492 and D492M.



**Figure 10.** Generation of D492M and D492HER2 from D492 in the EMP cell model.

The D492M and D492HER2 cell lines were generated from the D492 cell line. To generate D492M cells, D492 cells were cocultured with endothelial cells (BRENCs or HUVECs) to generate spindle colonies, followed by the subculture of the spindle colonies to generate the new stable cell line D492M, which is not tumorigenic. To generate D492HER2 cells, the HER2 (ERBB) receptor was overexpressed on D492 cells via transfection of the HER2-containing lentiviral plasmids, followed by a selection of the successfully transfected cells based on eGFP expression. In the end, the stable D492HER2 cell line was generated, which could form tumors in mice. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



**Figure 11.** The difference to the proteomes among the EMP cell lines.

Based on both the LFQ and SILAC proteomic analysis, there were 715 proteins significantly different between D492 and D492M, 535 proteins changed significantly from D492 to D492HER2, while 462 proteins were significantly different between D492M and D492HER2 (LFQ: Student's two-sample T-test, Permutation-based FDR < 0.05 & SILAC: Student's one-sample T-test, p value < 0.05). There were 312 proteins differently expressed in D492 compared with the other two cell lines (signature proteins for D492), and 97 differently expressed proteins in D492M (signature proteins for D492M), while 84 signature proteins in D492HER2 (LFQ: ANOVA, Permutation-based FDR < 0.05 & Coefficient of Variation of SILAC ratios < 0.1).

EMT is a heterogeneous process with as yet no universal criteria to define it. Groger et al. carried out a meta-analysis for EMT gene expression studies and identified an EMT-core gene list containing 130 dysregulated genes post-EMT (Gröger et al., 2012). Shaul and colleagues identified a list of metabolic markers in mesenchymal cell lines (MMS) that comprised 44 enzymes (Shaul et al., 2014). To characterize the EMP cell model in the framework of EMT, we compared the protein expression of the general and metabolic mesenchymal markers from the literature among D492, D492M, D492HER2 (**Figure 12A-B**). There was good consistency between LFQ and SILAC, with SILAC being more compatible with the literature. The EMT gene expression profile of the EMP model and that of the literature are concordant, apart from aldehyde dehydrogenase family 1 member A3 (ALDH1A3), 4-hydroxyphenylpyruvate dioxygenase-like protein (HPDL), aldo-keto reductase 1 member B1 (AKR1B1), and microsomal glutathione S-transferase 1 (MGST1).

ALDH1A3 and HPDL were reported to be decreased, while AKR1B1 and MGST1 were increased in the mesenchymal cells (Gröger et al., 2012; Shaul et al., 2014), which was contradictory to the EMP cell model. ALDH1A3 catalyzes the formation of retinoic acid and is important for embryonic development. HPDL localizes in mitochondria, and the biological function of which is less studied. AKR1B1 is an NADPH-dependent enzyme catalyzing the reduction of aldehydes and ketones. MGST1 regulates glutathione metabolism, oxidative stress, and cell detoxification. All four enzymes are reported to facilitate cancer progression (Duan et al., 2016; X. Wu et al., 2017; Ye et al., 2021; Zeng et al., 2020), and the discordance dampens their strength as reliable EMT markers.

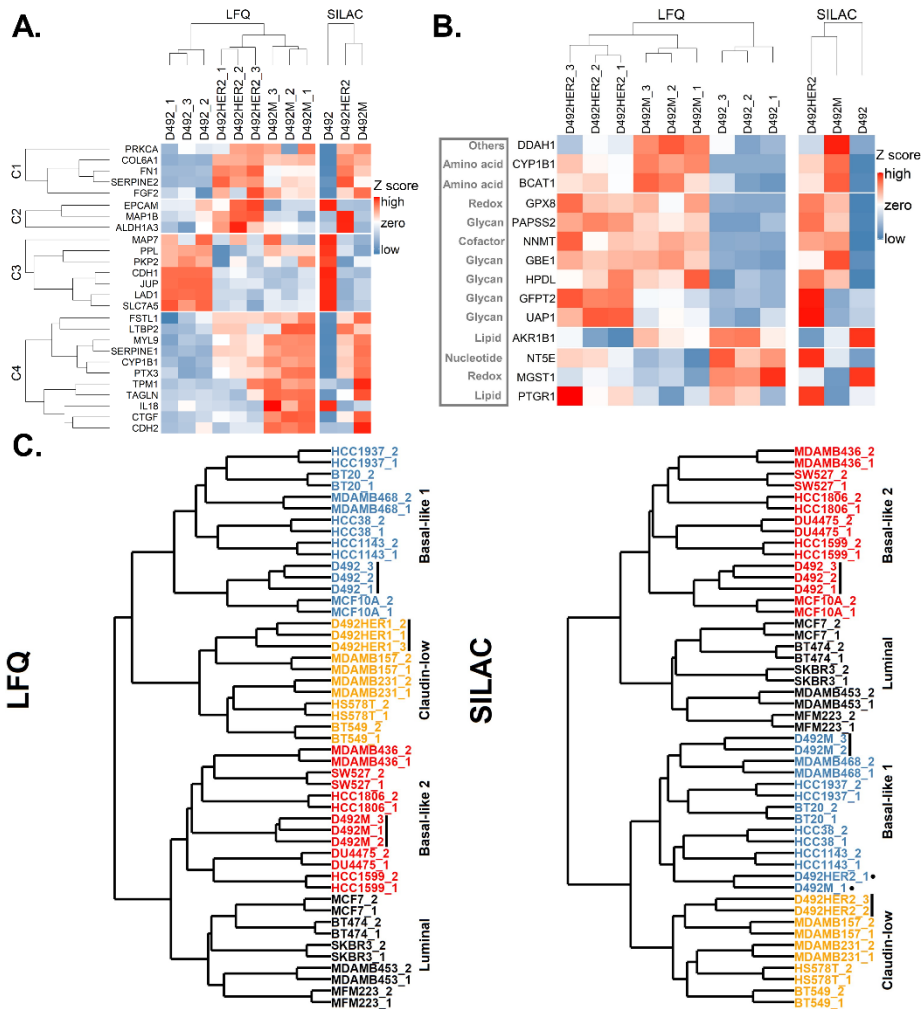
D492HER2 cell proteome is closer to that of D492M than D492, albeit unequivocal differences (**Figure 12A-B**). EMT marker expression in clusters C2 and C4 (**Figure 12A**) was discordant between D492HER2 and D492M, empathizing the accountabilities of these EMT markers in EMT plasticity.

Established on the reported proteomic fingerprints of 20 common breast epithelial cell lines (Lawrence et al., 2015), the unsupervised hierarchical clustering analysis revealed that D492 and D492M were basal-like breast cell lines, complying with the prior categorization of D492 (Sigurdsson et al., 2011), while D492HER2 was grouped with the “mesenchymal-like/claudin-low” cell lines based on both SILAC and LFQ (**Figure 12C**).

MCF10A cells express stem cell-like markers, and this cell line is non-tumorigenic and derived from normal breast tissue, which highly resembles

D492 (Bhat-Nakshatri et al., 2010; Neve et al., 2006; Qu et al., 2015; Soule et al., 1990). MDA-MB-468 is tumorigenic with metastatic capacity and was clustered with D492M possibly in consequence of its mesenchymal traits (Cailleau et al., 1978; H. J. Hugo et al., 2017; Neve et al., 2006). D492HER2 was clustered with the tumorigenic MDA-MB-157, originally isolated from metastatic human breast carcinoma (Cailleau et al., 1978; Neve et al., 2006; Young et al., 1974), and classified as a “mesenchymal-like/claudin-low” cell line based on both LFQ and SILAC. D492 and D492M imitate basal-like breast cells, while D492HER2 appears to be an intermediate phenotype with claudin-low properties diverting from D492 and D492M.

These findings strengthen the isogenic D492 cell lines as a valuable tool to study EMP and to define the molecular plasticity of EMT on the proteomic level.



**Figure 12.** Dysregulation of the general and metabolic EMT markers in the EMP cell model and cell line classification of D492, D492M, and D492HER2.

**(A)** Groger et al. was a published independent study on EMT marker dysregulation via a meta-analysis of gene expression studies (GES) of EMT, with a focus on different cell types and treatment modalities (Gröger et al., 2012). Based on this study, the identified EMT markers in the EMP cell model were confirmed. 26 out of the 130 EMT markers reported in the literature were detected in both the LFO and SILAC datasets. EPCAM and IL18 were inconsistent between LFO and SILAC, with the SILAC results being more similar to the literature. ALDH1A3 was inconsistent between the literature and this study. **(B)** Dysregulated metabolic EMT makers in the EMP model were compared with the MMS reported in Shaul et al., where the mRNA expression profiles of 1,704 metabolic genes in 978 human cancer cell lines were analyzed (Shaul et al., 2014). The LFO and SILAC quantification of all the metabolic EMT proteins agreed

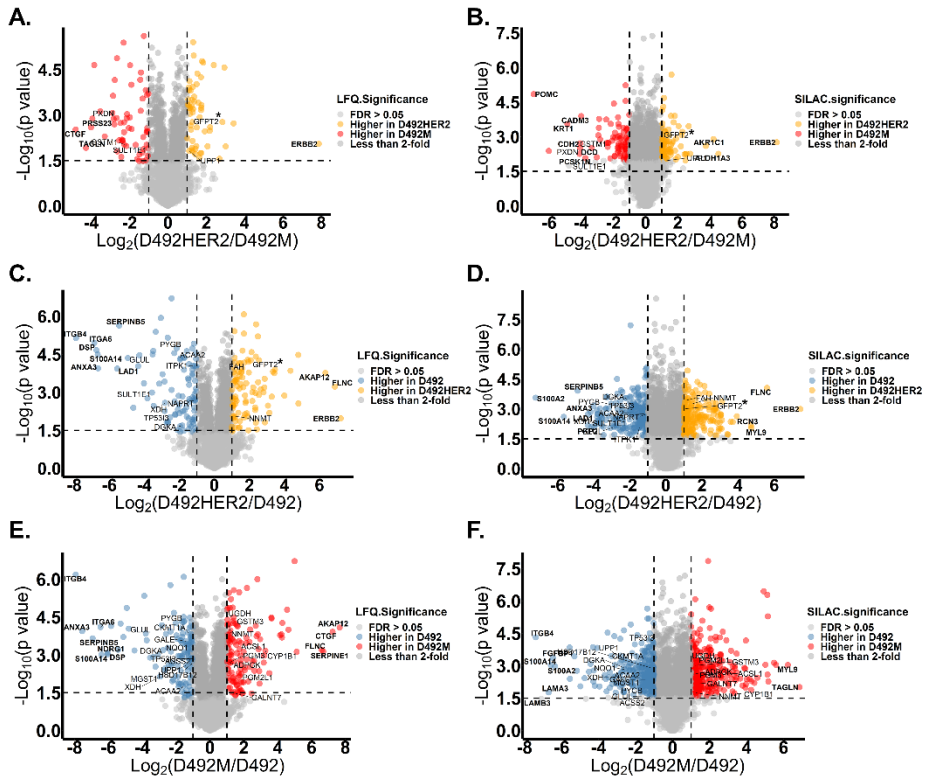
with each other except for NT5E, with SILAC being more consistent with the literature. Compared with the literature, HPDL, AKR1B1, and MGST1 altered oppositely in the EMP model. **(C)** Classification of the EMP cell model. Both the LFQ and SILAC raw data were re-quantified by the iBAQ quantification method in MaxQuant, and the expression of proteins identified in both literature and this study was used for the hierarchical clustering. D492, D492M, and D492HER2 were clustered with other pre-classified breast cell lines (Lawrence et al., 2015). In the literature, the breast epithelial cell lines commonly used in literature were classified into four groups: “luminal”, “basal-like 1,” “basal-like 2,” and “mesenchymal-like/claudin-low.” Based on the analysis, LFQ (left) classified D492 as “basal-like 1” (in blue), D492M as “basal-like 2” (in red), and D492HER2 as “mesenchymal-like/claudin-low” (in orange), while SILAC (right) classified D492 as “basal-like 2” (in red), D492M as “basal-like 1” (in blue), and D492HER2 as “mesenchymal-like/claudin-low” (in orange). Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

To characterize the proteomic changes post different types of EMT in the EMP cell model, supported by both LFQ and SILAC, we plotted the log<sub>2</sub> fold changes and -log<sub>10</sub> p values of all the detected proteins from both LFQ and SILAC and outlined the utmost changes among the three cell lines with arbitrary thresholds **(Figure 13)**. The topmost consistent non-metabolic change between D492HER2 and D492M was ERBB, owing to the transfection of the ERBB receptors into D492HER2 **(Figure 13A-B)**. FLNC, ERBB, SERPINB5, S100A14, ANXA3, and LAD1 were consistently altered between D492HER2 and D492 **(Figure 13C-D)**, while ITGB4 and S100A14 were outlined by the comparison between D492M and D492 **(Figure 13E-F)**.

The increased FLNC in the mesenchymal cells agrees with the involvement of FLNC in cell cytoskeleton modeling. SERPINB5 plummeted after EMT, supporting that SERPINB5 is an epithelial marker and competent for tumor suppression (Vecchi et al., 2008). ANXA3 has been reported to participate in chemoresistance (Tong et al., 2018), and the higher level of ANXA3 in D492 confirms the higher resistance of D492 to drug treatment (Barkovskaya et al., 2021). The ANXA3 expression in prostate cancer specimens is essentially reduced, making ANXA3 a prognostic marker in favor of patient survival (Köllermann et al., 2008). S100A14 has recently been reported to possess oncogenic effects in breast cancer (Xukun Li et al., 2020). It is drastically decreased in both D492HER2 and D492M, pinpointing its importance for D492. LAD1 is reported to be phosphorylated by EGF to promote tumor aggressiveness in breast cancer, which conflicts with the high level of LAD1 in D492 (Roth et al., 2018). S100A14 and LAD1 are reported oncogenes but significantly decreased in the mesenchymal cells, suggesting higher levels of S100A14 and LAD1 are not necessarily associated with higher carcinogenicity.



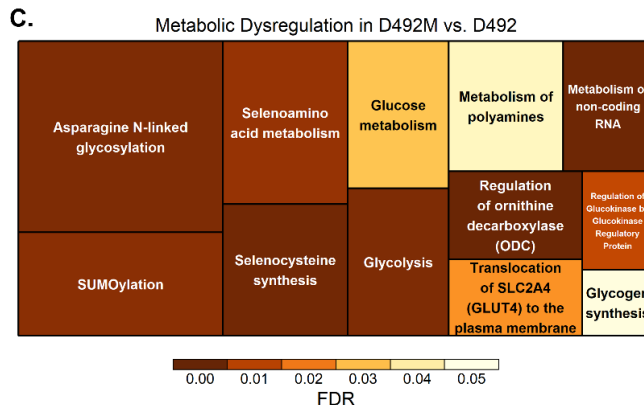
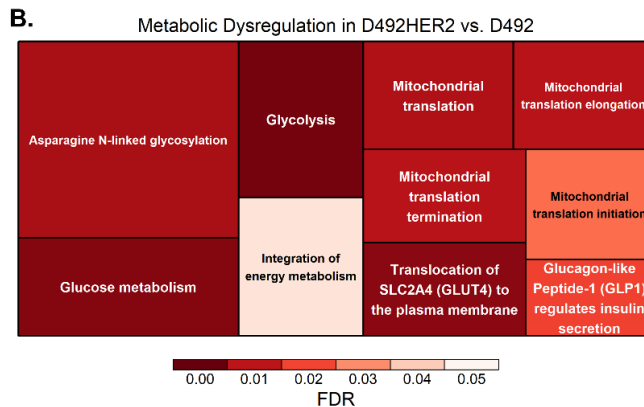
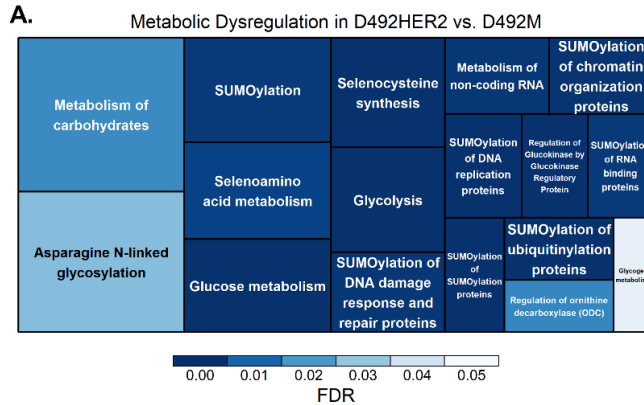
Reactome pathway analysis revealed that mitochondrial translation was primarily dysregulated in the partial EMT process represented by D492HER2 (**Figure 14B**), reflecting the significance of energy metabolism in cancer malignancy (Bergers and Fendt, 2021).



**Figure 13.** Proteomic changes among different cell lines in the EMP cell model were quantified by LFQ and SILAC.

Statistical analysis of the LFQ and SILAC proteome expression between two different cell lines in the EMP cell model: D492HER2 vs. D492M (**A-B**), D492 vs. D492HER2 (**C-D**), D492 vs. D492M (**E-F**). Proteins with significance and fold change > 2 were colored (LFQ: Student's T-test, two-sample tests, Permutation-based FDR < 0.05; SILAC: Student's T-test, one-sample tests, p value < 0.05). Metabolic enzymes were normally labeled for those that were differentially expressed in two cell lines and consistent between LFQ and SILAC. Proteins were boldly labeled for those that possessed significant difference and were with fold changes above certain thresholds (at least 4-fold difference between D492HER2 and D492M, 4-fold between D492HER2 and D492, and 6-fold between D492 and D492M). Horizontal dash lines indicated the  $-\log_{10}(p \text{ value})$  at 1.5, and vertical dash lines indicated the fold change at 2-fold. The top altered proteins between D492M and D492HER2 were PRSS23, CTGF, TAGLN, POMC, CADM3, KRT1, CDH2, DCD, PCSK1N, AKR1C1, ALDH1A3, and ERBB2, involved in cell adhesion and metabolism. The top differently expressed proteins in D492HER2 and D492 were AKAP12, FLNC, ERBB, RCN3, MYL9,

SERPINB5, ITGB4, ITGA6, DSP, S100A14, S100A2, LAD1, ANXA3, and PKP2, which were mainly involved in cell adhesion, structure, cell-cell interaction, and signaling. Similar to the difference observed with the other cell lines, the top differently altered proteins between D492 and D492M included AKAP12, CTGF, FLNC, SERPINE1, MYL9, TAGLN, ITGB4, ITGA6, ANXA3, SERPINB5, NDRG1, DSP, S100A14, FGFBP1, S100A2, LAMA3, and LAMB3. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



**Figure 14.** Functional changes among different cell lines in the EMP cell model.

Dysregulated Reactome metabolic pathways were enriched in D492HER2 vs. D492M **(A)**, D492HER2 vs. D492 **(B)**, and D492M vs. D492 **(C)**. Proteins significantly changed between two cell lines in both LFQ and SILAC were included for the Reactome pathway analysis. Student's T-test, Permutation-based FDR < 0.05, one sample T-test, p value of SILAC ratio < 0.05. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

## 4.2 Search for breast EMT markers

In this section, I described the consistent EMT markers identified in the essential spontaneous EMT program among the three EMT cell models. Moreover, the proteomic profile of the partial mesenchymal D492HER2 in the EMP cell model was explored by comparing the proteome of D492HER2 with that of the epithelial D492 and mesenchymal D492M.

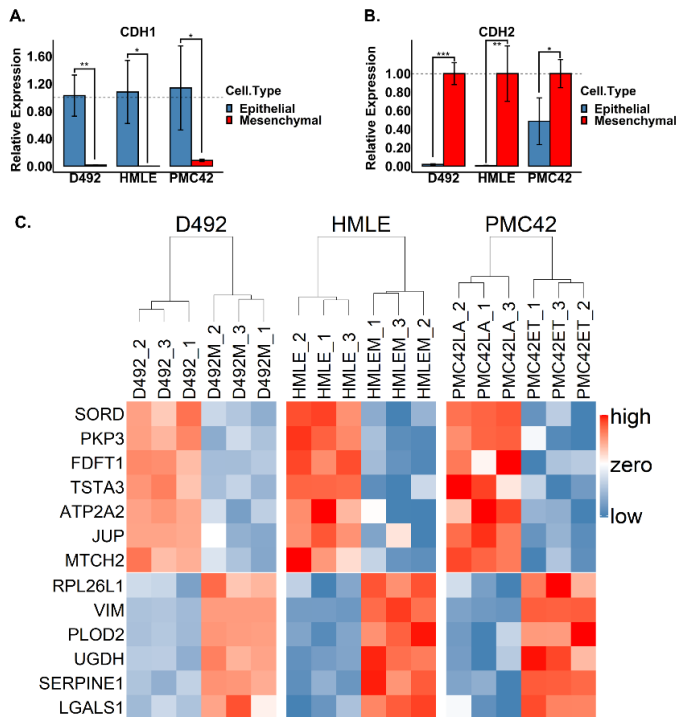
### 4.2.1 Common EMT markers were identified (*paper I*)

Heterogeneity and plasticity are two of the main molecular properties of the EMT program, which encourages the field to identify a set of consistent and reliable markers to define EMT. We identified 13 consistently changed proteins post-EMT, within which 4 metabolic enzymes were discovered.

One hallmark of EMT is the switch between *E*-cadherin (CDH1) and *N*-cadherin (CDH2) expression (Loh et al., 2019). Forasmuch as the proteomic data were unable to detect CDH1 and CDH2, real-time PCR was applied for probing the RNA expression of these two EMT markers. The downregulated CDH1 and upregulated CDH2 in all three EMT cell models confirmed the *E*-cadherin and *N*-cadherin switch in EMT **(Figure 15A-B)**.

The consistently dysregulated proteins after EMT in all three EMT cell models were summarized **(Figure 15C)**. The upregulated VIM, LGALS1, and SERPINE1 and downregulated PKP3 in the mesenchymal cells were EMT markers recorded in a public EMT database (M. Zhao et al., 2019). A thorough literature review for the identified EMT targets revealed that all had been connected to EMT, albeit to different extents, with many pending to be explored in the context of EMT **(Table 5)**. The regulation of these EMT markers was further confirmed in the tumorigenic breast mesenchymal cell line D492HER2 in the context of cancer progression, with the same trends in changes seen for all the targets detected in D492HER2 **(Table 5)**. It indicates that these markers are not responsible for the tumor initiation but are critical for EMT and are potentially involved in cancer malignancy.

VIM, SERPINE1, LGALS1, and PKP3 have been related to EMT in different cancer types (Aigner et al., 2007; Bacigalupo et al., 2015; Bedi et al., 2014; Hui Li et al., 2017; Pavón et al., 2012; F. Xu et al., 2017; Yamagami et al., 2020; J. Yang et al., 2020). Vimentin, a type III intermediate filament and a well-known EMT marker, shapes cell structure and modifies cell movement and adhesion (Mendez et al., 2010). SERPINE1, a key player in endothelial homeostasis, is highly upregulated in EMT. However, the function of SERPINE 1 in EMT is poorly understood. The possible role of SERPINE1 in EMT is to affect the function of urokinase-type plasminogen activator receptor (uPAR) to regulate ECM degradation (Yamagami et al., 2020). LGALS1 is a carbohydrate-binding protein. A study showed that upregulation of LGALS1 decreased CDH1 and increased SNAI1 (Bacigalupo et al., 2015). PKP3 is an epithelial marker and is under the control of EMT transcription regulator ZEB1 (Aigner et al., 2007; Bedi et al., 2014).



**Figure 15.** Consistently altered proteins post-EMT in all three EMT cell models.

**(A-B)** The RNA expression of CDH1 was downregulated **(A)**, while that of CDH2 was upregulated after EMT in all three EMT cell models **(B)**. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . **(C)** The proteomic analysis identified 13 proteins significantly changed (Student's T-test,  $p < 0.05$ ) in all three EMT models and with the same up- or down-regulated directions. Within these proteins, SERPINE1, RPL26L1, PLOD2, UGDH, LGALS1, and VIM were upregulated, while JUP, PKP3, MTCH2,

ATP2A2, FDFT1, SORD, and TSTA3 were downregulated after EMT. UGDH, FDFT1, SORD, and TSTA3 were metabolic proteins. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

#### 4.2.2 Distinct proteomic signatures of the EMP cell line (*paper II*)

Accordingly, D492HER2 resides at the EMT spectrum created by the epithelial D492 and mesenchymal D492M, with increased invasiveness. Identifying the altered proteome in D492HER2 compared with the other cell lines (signature proteins) advances the searching for accountable markers in the EMT-involved oncogenicity. 65 out of the 84 identified dysregulated proteins in D492HER2 were upregulated, implying increased cellular activities (**Table 6**). The proteomic signature consisted of a broad range of proteins with diverse cellular functions, covering DNA replication, RNA transcription, protein translation, cell structure and movement, transport, cell signaling, and metabolism.

Focusing on the distinctive metabolic phenotype in D492HER2, GFPT2, SLC16A3, ALDH9A1, HK1, MGEA5, ACLY, and NDUFAB1 were upregulated, while cytosolic FH, SOD2, and GLA were downregulated. SLC16A3 encodes monocarboxylate transporters, carrying lactate, keto bodies, and more across the plasma membrane, and is a marker for increased dependence for glycolysis (Halestrap, 2013). ALDH9A1 belongs to aldehyde dehydrogenases, correlated with cell proliferation (Muzio et al., 2012). HK1 participates in glucose metabolism. Cytosolic FH has been shown as a tumor suppressor and part of the DNA damage response (Yogev et al., 2010). GFPT2 that generates UDP-GlcNAc, MGEA5 that removes O-GlcNAc, and GLA that hydrolyzes glycolipids are enzymes related to glycans. ACLY is a lipogenic enzyme critical for the *de novo* lipid synthesis by converting cytoplasmic citrate to acetyl-CoA and oxaloacetate and is a key player in cancer metabolism (Zaidi et al., 2012). NDUFAB1 is also known as acyl carrier protein (ACP), takes part in the type II fatty acid synthesis in mitochondria, and has been shown to regulate mitochondrial bioenergetics and ROS metabolism to affect cell viability (Feng et al., 2009; T. Hou et al., 2019). SOD2 is an oxidant scavenger in mitochondria (Idelchik et al., 2017) and is one of the few enzymes downregulated in D492HER2, possibly owning the impaired mitochondria.

Taken together, D492HER2 regulates many oncogenic enzymes required for cell proliferation, lipid synthesis, glycan metabolism, mitochondria and ROS regulation, DNA damage response, and central carbon metabolism, rendering it more invasiveness than the other two EMP cell lines.

**Table 5.** Consistent proteomic changes in EMT.

The expression of consistently changed proteins among the three EMT cell models was reported. The EMT targets were significantly different (Student's T-test,  $p < 0.05$ ). The literature related to each target in terms of EMT was also listed. Changes of these targets in another breast mesenchymal cell line with tumorigenicity (D492HER2) were consistent with the findings in this study. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

Uniprot ID	Protein Name	Gene Name	Log2(D492/D492M)	Log2(HMLE/HMLEM)	Log2(PMC42LA/PMC42ET)	-Log(p-value D492s)	-Log(p-value HMLEs)	-Log(p-value PMC42s)	EMT-related Literature	Gene expression in mesenchymal cells with tumorigenicity (Log2Ratio)
P05121-1	Plasminogen activator inhibitor 1	SERPINE1	-4.236	-3.442	-3.993	3.416	3.313	3.780	(Pavón et al., 2012; Xu et al., 2019; Yamagami et al., 2020)	-4.739
Q9UNX3	60S ribosomal protein L26-like 1	RPL26L1	-2.367	-1.123	-1.346	1.981	2.349	1.925	(Piskareva et al., 2015)	Not detected
O00469-1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	-1.817	-1.504	-1.121	5.005	3.411	2.072	(Pavón et al., 2012; Xu et al., 2017)	-1.687
O60701	UDP-glucose 6-dehydrogenase	UGDH	-1.193	-2.550	-0.779	3.101	3.962	2.285	(Arnold et al., 2019; Vergara et al., 2015; Lin et al., 2020; Teoh et al., 2020; Wang et al., 2019)	-0.828
P09382	Galectin-1	LGALS1	-0.623	-2.912	-0.444	1.769	3.035	1.799	(Bacigalupo et al., 2015; Li et al., 2017; Pavón et al., 2012)	-0.422
P08670	Vimentin	VIM	-0.407	-1.979	-3.343	5.493	4.802	5.970	(Yang et al., 2020)	-0.957
Q13630	GDP-L-fucose synthase	TSTA3	0.553	1.009	0.508	2.871	2.517	1.789	(Ren et al., 2015; Wang et al., 2020)	0.562
Q00796	Sorbitol dehydrogenase	SORD	0.599	1.459	1.452	2.457	2.957	2.806	(Schwab et al., 2018)	0.111
P37268	squalene synthase	FDFT1	0.683	1.536	0.491	3.510	3.215	1.462	(Yang et al., 2020b)	0.722
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	0.919	0.523	0.641	2.388	1.551	2.316	(Davis et al., 2013)	0.299
Q9Y6C9	Mitochondrial carrier homolog 2	MTCH2	1.239	0.725	0.899	2.219	1.624	3.015	(Ren et al., 2015)	0.198
Q9Y446	Plakophilin-3	PKP3	2.958	2.057	3.478	2.786	3.077	1.850	(Aigner et al., 2007; Bedi et al., 2014)	3.458
P14923	Junction plakoglobin	JUP	4.343	1.541	1.214	1.783	1.356	3.611	(Demirkan, 2013; Lourenco et al., 2020)	4.344

**Table 6.** Signature proteins in partial mesenchymal D492HER2.

Proteins were significantly down-/up- regulated in D492HER2 compared with the other two cell types (D492 and D492M) based on both LFQ and SILAC. The z-score values of the protein expression in D492HER2 were reported. Data used were from (Q. Wang et al., 2021).

Protein ID	Protein Name	Gene Name	LFQ_D492HER2	<sup>a</sup> ANOVA.FDR_D492HER2	SILAC_D492HER2	<sup>b</sup> CV_D492HER2
Q9NR33	DNA polymerase epsilon subunit 4	POLE4	-2.123	0.006	-1.133	0.064
Q53Y83	Galactosidase, alpha	GLA	-2.031	0.012	-1.152	0.065
Q5VXV3	SET	SET	-1.828	0.002	-1.150	0.048
Q95758-1	Isoform 1 of Polypyrimidine tract-binding protein 3	PTBP3	-1.629	0.001	-1.154	0.003
A0A024R829	Polymerase (DNA directed), epsilon 3 (P17 subunit), isoform CRA_a	POLE3	-1.553	0.002	-1.152	0.032
Q9BPX3	Condensin complex subunit 3	NCAPG	-1.471	0.009	-1.145	0.001
Q9NQG5	Regulation of nuclear pre-mRNA domain-containing protein 1B	RPRD1B	-1.390	0.001	-1.140	0.013
A0A024R2Q3	Catenin (Cadherin-associated protein), beta 1, 88kDa, isoform CRA_a	CTNNB1	-1.370	0.001	-1.150	0.066
P19105	Myosin regulatory light chain 12A	MYL12A	-1.219	0.000	-1.148	0.040
Q5VXN0	Ribosome production factor 2 homolog (Fragment)	RPF2	-1.058	0.012	-1.147	0.079
A0A024R694	Actinin, alpha 1, isoform CRA_a	ACTN1	-0.920	0.002	-1.142	0.070
Q75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	-0.850	0.001	-1.155	0.021
P04179-4	Isoform 4 of Superoxide dismutase [Mn], mitochondrial	SOD2	-0.828	0.030	-1.137	0.013
A0A024R395	MRE11 meiotic recombination 11 homolog A (S. cerevisiae), isoform CRA_a	MRE11A	-0.781	0.000	-1.136	0.081
Q53Y51	D-dopachrome tautomerase	DDT	-0.776	0.017	-1.154	0.040
P40616-2	Isoform 2 of ADP-ribosylation factor-like protein 1	ARL1	-0.717	0.042	-1.142	0.050
Q6FGX3	RAB6A protein	RAB6A	-0.355	0.009	-1.146	0.009
P38159	RNA-binding motif protein, X chromosome	RBMX	-0.216	0.008	-1.133	0.053
P07954-2	Isoform Cytoplasmic of Fumarate hydratase, mitochondrial	FH	-0.164	0.037	-1.143	0.034
Q9LUPN1	Serine/threonine-protein phosphatase (Fragment)	PPP1CC	0.098	0.004	1.155	0.060
P62070	Ras-related protein R-Ras2	RRAS2	0.118	0.005	1.078	0.088
Q9NXH9-2	Isoform 2 of tRNA (guanine(26)-N(2))-dimethyltransferase	TRMT1	0.129	0.002	1.136	0.011
Q53T99	Ribosome biogenesis protein WDR12	WDR12	0.199	0.039	1.153	0.084
C9J2Y9	DNA-directed RNA polymerase subunit beta	POLR2B	0.307	0.002	0.841	0.006
Q7Z471	Nuclear fragile X mental retardation-interacting protein 2	NUFIP2	0.362	0.013	1.119	0.073
Q7Z3B4-3	Isoform 3 of Nucleoporin p54	NUP54	0.371	0.010	1.134	0.067
J3KNL6	Protein transport protein sec16	SEC16A	0.468	0.003	1.154	0.035
Q06830	Peroxiredoxin-1	PRDX1	0.502	0.003	1.086	0.022
Q61AX5	Eukaryotic translation initiation factor 3 subunit E	EIF3E	0.525	0.004	1.002	0.066
Q8WXF1	Paraspeckle component 1	PSPC1	0.562	0.012	1.142	0.078

**Footnotes:**

a: in the LFQ experiment, the statistical analysis was ANOVA, and proteins with FDR < 0.05 were listed.

b: in the SILAC experiment, proteins with Coefficient of Variation (CV) of SILAC ratios < 0.1 were listed.

Table 6. (Continued)

Protein ID	Protein Name	Gene Name	LFQ_D492HER2	<sup>a</sup> ANOVA_FDR_D492HER2	SILAC_D492HER2	<sup>b</sup> CV_D492HER2
B4DVQ5	Eukaryotic translation initiation factor 3 subunit C	EIF3C	0.573	0.001	1.116	0.085
Q0VGA5	SARS protein	SARS	0.577	0.010	1.116	0.003
O15371-2	Isoform 2 of Eukaryotic translation initiation factor 3 subunit D	EIF3D	0.622	0.003	0.972	0.079
Q9Y262	Eukaryotic translation initiation factor 3 subunit L	EIF3L	0.623	0.007	1.071	0.068
P78406	mRNA export factor	RAE1	0.630	0.001	1.059	0.060
Q7Z7E8	Ubiquitin-conjugating enzyme E2 Q1	UBE2Q1	0.635	0.038	1.034	0.046
O14561	Acyl carrier protein, mitochondrial	NDUFAB1	0.720	0.010	1.128	0.036
P55795	Heterogeneous nuclear ribonucleoprotein H2	HNRNPH2	0.779	0.021	1.014	0.088
A0A024RCR5	Bromodomain containing 2, isoform CRA_b	BRD2	0.779	0.044	1.110	0.063
P42285	Superkiller viralicidal activity 2-like 2	SKIV2L2	0.817	0.000	1.143	0.051
Q6FHL9	PEA15 protein	PEA15	0.872	0.001	1.154	0.017
A0A024R370	TATA element modulatory factor 1, isoform CRA_a	TMF1	0.898	0.011	1.133	0.021
Q14157-4	Isoform 4 of Ubiquitin-associated protein 2-like	UBAP2L	0.907	0.005	0.942	0.052
A0A024R1T9	ATP-citrate synthase	ACLY	0.914	0.000	1.108	0.060
Q9UKL0	REST corepressor 1	RCOR1	0.946	0.023	0.913	0.038
A0A140VJJ2	S-formylglutathione hydrolase	ESD	0.953	0.001	1.154	0.038
A0A024R9T6	HCG17415, isoform CRA_a	SLAIN2	0.955	0.015	1.118	0.067
Q95490-2	Isoform 2 of Adhesion G protein-coupled receptor L2	ADGRL2	1.018	0.000	1.138	0.046
A0A024R6K8	Tryptophanyl-tRNA synthetase, isoform CRA_a	WARS	1.025	0.020	1.148	0.042
Q5JTV8-3	Isoform 3 of Torsin-1A-interacting protein 1	TOR1AIP1	1.029	0.001	1.029	0.019
O60502	Protein O-GlcNAcase	MGEA5	1.036	0.006	1.141	0.057
A0A0S2Z3H8	GNAS complex locus isoform 1 (Fragment)	GNAS	1.040	0.002	1.124	0.051
A0A024RDU9	General transcription factor IIF subunit 2	GTF2F2	1.063	0.024	1.108	0.062
A0A024RDR0	High-mobility group box 1, isoform CRA_a	HMGB1	1.066	0.001	1.126	0.044
Q9BV57	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	ADI1	1.071	0.002	1.146	0.088
Q9Y678	Coatomer subunit gamma-1	COPG1	1.073	0.000	0.814	0.085
Q14566	DNA replication licensing factor MCM6	MCM6	1.173	0.003	1.127	0.051
Q9NQC3-5	Isoform 5 of Reticulon-4	RTN4	1.194	0.006	1.024	0.073
Q9H7D7-2	Isoform 2 of WD repeat-containing protein 26	WDR26	1.200	0.001	0.980	0.030
P49792	E3 SUMO-protein ligase RanBP2	RANBP2	1.201	0.001	1.092	0.052
Q96CP5	PMPCB protein (Fragment)	PMPCB	1.204	0.003	1.134	0.003
Q14008	Cytoskeleton-associated protein 5	CKAP5	1.265	0.000	0.822	0.015

## Footnotes:

a: in the LFQ experiment, the statistical analysis was ANOVA, and proteins with FDR < 0.05 were listed.

b: in the SILAC experiment, proteins with Coefficient of Variation (CV) of SILAC ratios < 0.1 were listed.



Table 6. (Continued)

Protein ID	Protein Name	Gene Name	LFQ_D492HER2	<sup>a</sup> ANOVA.FDR_D492HER2	SILAC_D492HER2	<sup>b</sup> CV_D492HER2
Q9NZT2-2	Isoform 2 of Opioid growth factor receptor	OGFR	1.266	0.000	1.029	0.010
A0A024R4G1	Leucine rich repeat containing 47, isoform CRA_a	LRRC47	1.308	0.001	1.121	0.035
Q9C0C9	(E3-independent) E2 ubiquitin-conjugating enzyme	UBE2O	1.313	0.014	0.803	0.023
Q9NVI7-2	Isoform 2 of ATPase family AAA domain-containing protein 3A	ATAD3A	1.335	0.000	1.003	0.035
P30085	UMP-CMP kinase	CMPK1	1.365	0.001	1.030	0.042
A0A024RC61	Aminopeptidase	ANPEP	1.395	0.000	1.131	0.058
P19367-4	Isoform 4 of Hexokinase-1	HK1	1.397	0.001	1.092	0.063
P35606-2	Isoform 2 of Coatomer subunit beta	COPB2	1.399	0.001	1.056	0.028
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	GFPT2	1.441	0.000	1.146	0.052
O60664-4	Isoform 4 of Perilipin-3	PLIN3	1.479	0.000	1.154	0.009
A0A024R313	Glycosyltransferase 8 domain containing 1, isoform CRA_a	GLT8D1	1.506	0.006	1.126	0.009
Q5H9A7	Metalloproteinase inhibitor 1	TIMP1	1.531	0.002	1.109	0.092
O43493-2	Isoform TGN46 of Trans-Golgi network integral membrane protein 2	TGOLN2	1.563	0.001	1.146	0.032
Q9UMS4	Pre-mRNA-processing factor 19	PRPF19	1.587	0.019	1.146	0.024
E7EVA0	Microtubule-associated protein	MAP4	1.649	0.001	1.155	0.020
P49189	4-trimethylaminobutylaldehyde dehydrogenase	ALDH9A1	1.649	0.001	0.894	0.012
A0A024R8U1	Solute carrier family 16 (Monocarboxylic acid transporters), member 3, isoform CR/	SLC16A3	1.791	0.002	0.825	0.033
A0A024QZW7	Nucleoporin 153kDa, isoform CRA_a	NUP153	1.800	0.004	1.142	0.035
Q9NVD7	Alpha-parvin	PARVA	1.847	0.039	1.126	0.076
P53621	Coatomer subunit alpha	COPA	1.881	0.001	0.956	0.046
Q01433-3	Isoform Ex1A-3 of AMP deaminase 2	AMPD2	1.928	0.002	1.001	0.017
O14579	Coatomer subunit epsilon	COPE	2.089	0.004	1.108	0.012

## Footnotes:

a: in the LFQ experiment, the statistical analysis was ANOVA, and proteins with FDR < 0.05 were listed.

b: in the SILAC experiment, proteins with Coefficient of Variation (CV) of SILAC ratios < 0.1 were listed.

### 4.3 Metabolic alteration in breast EMT

Herein we set out to explore the uniformly changed metabolic proteins in the EMT cell models in breast epithelium. In addition, the abnormally regulated enzymes among the EMP breast cell lines were also depicted.

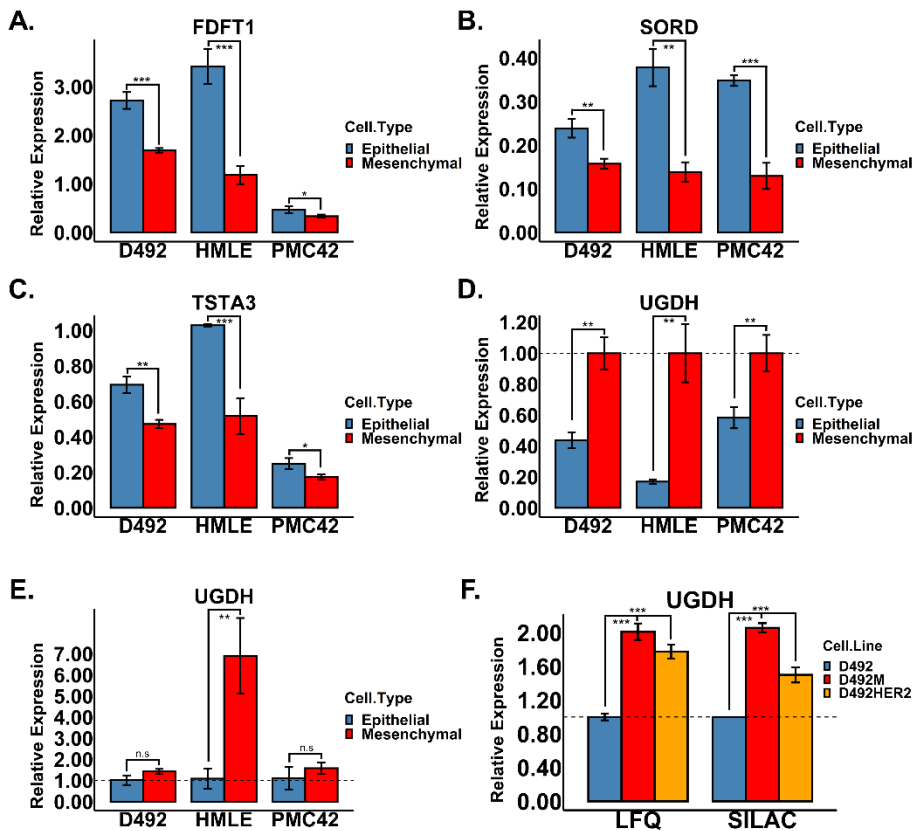
#### 4.3.1 The commonly altered metabolic targets post-EMT (*paper I*)

The metabolic enzymes FDFT1, SORD, and TSTA3 were significantly downregulated, while UGDH was upregulated and altered the most (**Figure 16A-D**). The upregulation of UGDH after EMT was confirmed on the RNA level and further validated in the context of breast cancer progression. The increasing trends in the UGDH RNA expression after EMT were observed in all EMT cell models (**Figure 16E**), and the protein expression of UGDH was higher in the tumorigenic breast mesenchymal cell line D492HER2 compared with D492, consistent with the non-tumorigenic D492M (**Figure 16F**).

The EMT-associated functions of the metabolic targets FDFT1, SORD, and TSTA3 have not been well investigated in literature (**Figure 16**). FDFT1, or squalene synthase, also short for SQR, is involved in cholesterol metabolism and has been related to cancer metastasis (Ha and Lee, 2020). FDFT1 is marginally associated with the survival of patients with basal breast cancer, according to kmplot.com. However, it has not been associated with EMT. SORD oxidizes sorbitol to fructose in the polyol pathway to metabolize the excess glucose, and a study has shown that a stable knockdown of SORD could block EMT (Schwab et al., 2018). This observation is contradictory to what has been seen in this study, where SORD is downregulated in the mesenchymal cells. TSTA3, also known as GFUS, catalyzes the production of GDP-fucose, which is an essential substrate for fucosylation. TSTA3 has been proposed as an oncogenic target, and high expression of TSTA3 is correlated with poor survival in patients with breast cancer based on kmplot.com and esophageal squamous cell carcinoma (Sun et al., 2016; J. Yang et al., 2018). Since TSTA3 is downregulated in EMT, this suggests that the poor survival rate in patients with high TSTA3 expression is not related to EMT.

The well-known functions of UGDH include production of hyaluronan, glucuronidation of xenobiotics, synthesis of proteoglycans, and influences on protein glycosylation. Recently, many studies have reported that UGDH is involved in cancer progression and participates in tumor growth, metastasis, and patient survival (Arnold et al., 2019; Goodwin et al., 2019; Huang et al.,

2010; Lin et al., 2020; Oyinlade et al., 2018; Teoh et al., 2020). UGDH has been associated with EMT; however, the exact role of UGDH in EMT is complex and inexplicit (Arnold et al., 2019; Lin et al., 2020; Teoh et al., 2020; Vergara et al., 2015; X. Wang et al., 2019). Arnold et al. has recently found that UGDH is a clinically relevant metabolic enzyme that is highly expressed in mesenchymal-like breast cancers (Arnold et al., 2019). Of late, Wang and colleagues published on *NATURE LETTER*, reporting that UGDH controls the availability of UDP-Glc to regulate the main EMT transcription factor SNAI1 (X. Wang et al., 2019). The upregulation of UGDH in both D492M and D492HER2 (**Figure 16F**) suggests that UGDH engages in the mesenchymal characteristics in cancer invasiveness.



**Figure 16.** Consistently altered metabolic proteins post-EMT in all three EMT cell models with UGDH altered the most.

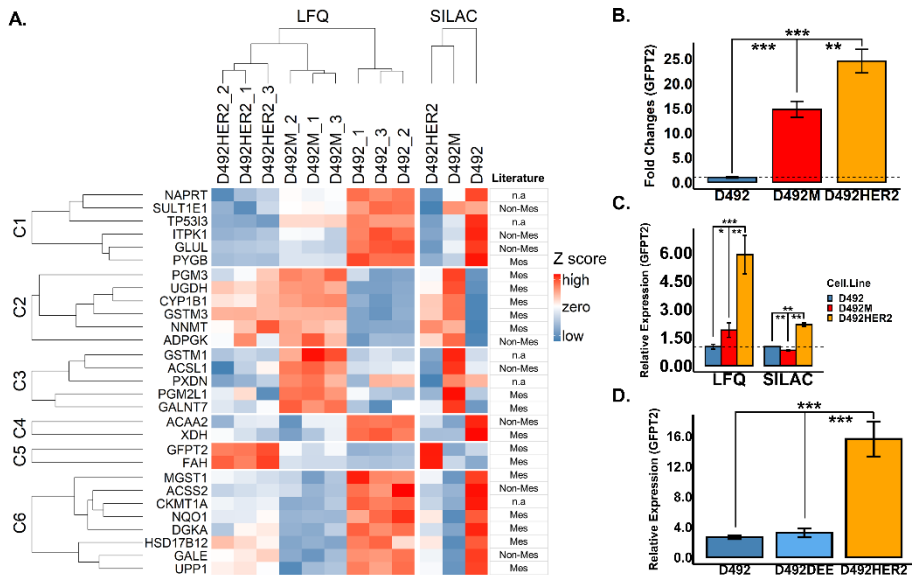
(A-D) The metabolic targets, i.e., FDFT1, SORD, TSTA3, and UGDH, were identified in the proteomic analysis and changed consistently in all three EMT cell models. (E) The mesenchymal cell lines in all three EMT cell models consistently expressed a higher level of the UGDH RNA. (F) The UGDH protein levels in D492 and D492M

were confirmed in the EMP datasets and further confirmed in the tumorigenic breast mesenchymal cell line D492HER2. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . FDFT1: farnesyl-diphosphate farnesyltransferase 1; SORD: sorbitol dehydrogenase; TSTA3: GDP-L-fucose synthetase; UGDH: UDP-glucose dehydrogenase. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

### 4.3.2 Different metabolic profiles of the EMP cell model (*paper II*)

In this chapter, we identified metabolic proteins with more than two-fold changes in expression between two types of cell lines, and these changes were confirmed by both LFQ and SILAC with a pre-defined significance threshold (**Figure 17A**). We grouped the identified enzymes into 6 clusters on account of the six possible comparison profiles among the three cell lines. These metabolic targets extended several KEGG metabolic pathways with several targets previously connected to EMT (**Table 7**). Most enzymes were associated with the metabolism of glycan precursors, i.e., PYGB, PGM3, UGDH, PGM2L1, GALNT7, GFPT2, and GALE, which were indiscriminately spanned among different clusters. GFPT2 and FAH were solely highly expressed in the partial mesenchymal D492HER2, with GFPT2 being more significant. The highest RNA expression of GFPT2 was detected in D492HER2 (**Figure 17B**), supporting the proteomic result (**Figure 17C**).

FAH is a key enzyme of the tyrosine catabolic pathway and is widely studied in tyrosinemia (M. Zhu et al., 2019). GFPT2 catalyzes fructose-6-phosphate into glucosamine-6-phosphate, concurrently converting glutamine into glutamate, and is the rate-limiting enzyme in the HBP. The end-product of the HBP is UDP-GlcNAc, an indispensable nucleotide sugar that partakes in protein glycosylation. Various transcription factors, signaling mediators, and metabolic enzymes are modulated by O-GlcNAcylation in cancer, and glycans are critical for ECM modeling in a malignant environment (Chiaradonna et al., 2018; K. Taparra et al., 2016). Protein GlcNAcylation facilitates breast cancer metastasis and tumorigenesis (Y. Gu et al., 2010). The function and regulation of GFPT2 in protein O-GlcNAcylation have drawn a considerable amount of attention by virtue of the UDP-GlcNAc production by GFPT2 (L. Liu et al., 2020; Szymura et al., 2019; Zhou et al., 2019). GFPT2 is one of the MMS genes identified by Shaul and colleagues (Shaul et al., 2014) and is associated with the mesenchymal phenotypes of invasive breast cancer (Simpson et al., 2012). It has also been reported that GFPT2 engages in regulating cellular oxidative stress, the mechanism of which remains yet unclear, likely owing to the glutaminolysis mediated by GFPT2 (Askari et al., 2019; Chao et al., 2021; Zitzler et al., 2004). We concentrated on GFPT2 as the key mesenchymal feature in cancer progression.



**Figure 17.** Altered metabolic proteins post two types of EMT processes in the EMP cell model with GFPT2 altered the most in D492HER2 compared with the other two cell lines.

**(A)** Metabolic proteins that expressed differentially in D492HER2 vs. D492M, D492HER2 vs. D492, and D492M vs. D492 were manually identified and plotted (Student's T-test, permutation-based FDR < 0.05 for LFQ, one sample T-test, p value of SILAC < 0.05, more than 2-fold in both LFQ and SILAC). The median relative SILAC expression of each target for D492, D492M, and D492HER2 was used for plotting. The metabolic targets were manually clustered into six clusters. The relative expression of each metabolic protein was indicated in the color bar, scaling from blue to red, representing the lowest to the highest expression. The identified metabolic targets were classified into two groups based on the literature (Shaul et al., 2014): Mesenchymal (Mes) and non-mesenchymal (Non-Mes). n.a: not available in the literature. **(B-D)** The expression of GFPT2 in the EMP cell model. On the RNA level **(B)**, GFPT2 expressed at the highest level in D492HER2, while it had the lowest expression in D492. **(C)** The GFPT2 protein expression was the highest in D492HER2 confirmed by both LFQ (left) and SILAC (right), followed by D492M, with D492 expressed the lowest. **(D)** D492DEE was the negative control cell line of D492HER2 for HER2 overexpression, the low GFPT2 level in D492DEE indicated that the increased expression of GFPT2 in D492HER2 was not due to the cell handling process in HER2 induction. Student's T-test, \*: p < 0.05; \*\*: p < 0.01; \*\*\*: p < 0.001. GFPT2: glutamine-fructose-6-phosphate transaminase 2; LFQ: label-free quantification; SILAC: stable isotope labeling by amino acids in cell culture. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

**Table 7.** Main metabolic targets dysregulated in the EMP cell model. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

The significantly altered metabolic targets were listed by comparing D492HER2 with D492M, D492HER2 with D492, and D492M with D492 (Permutation-based FDR < 0.05) with at least 2-fold difference, confirmed by LFQ and SILAC. The reported log2 ratios were the average of the LFQ and SILAC ratios. The relevant transcription factors and literature were reported for each target based on literature review.

Protein ID	Protein Names	Gene Name	KEGG Classification	Log2(D492HER2/D492M)	Possible Transcription Factors	Citation related to EMT
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	GFPT2	Carbohydrate metabolism	1.558	NF-KB; SIRT6; BMP-2	(Shaul et al., 2014; Simpson et al., 2012; Szymura et al., 2019; Taparra et al., 2019; Zhang et al., 2018; Zhou et al., 2019)
Q16831	Uridine phosphorylase 1	UPP1	Pyrimidine metabolism	1.167	NF-Kb; Oct3/4	(Guan et al., 2019; Wehbe et al., 2012)
X5DR03	Glutathione S-transferase mu 1 isoform B (Fragment)	GSTM1	Glutathione metabolism	-2.152	Nrf2	n.a
Q53X91	Sulfotransferase (Fragment)	SULT1E1	Steroid hormone biosynthesis	-2.383	Nrf2	n.a
Q92626	Peroxidase homolog	PXDN	Oxidoreductases	-3.825	Snai1; Nrf2	(Briem et al., 2019; Sitole and Mavi-Damelin, 2018)
<b>Log2(D492HER2/D492)</b>						
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	GFPT2	Carbohydrate metabolism	1.827	NF-KB; SIRT6; BMP-2	(Shaul et al., 2014; Simpson et al., 2012; Szymura et al., 2019; Taparra et al., 2019; Zhang et al., 2018; Zhou et al., 2019)
Q6FH49	NNMT protein	NNMT	Nicotinate and nicotinamide metabolism	1.275	Stat3	(Eckert et al., 2019; Shaul et al., 2014)
P16930	Fumarylacetoacetase	FAH	Tyrosine metabolism	1.272	CDC5L	n.a
Q6XQN6	Nicotinate phosphoribosyltransferase	NAPRT	Nicotinate and nicotinamide metabolism	-1.266	NF-Kb; STAT3; HIF-1a	(Lee et al., 2018)
AA0A024R6H3	Inositol 1,3,4-triphosphate 5/6 kinase, isoform CRA_a	ITPK1	Inositol phosphate metabolism	-1.358	BMP2; TBX2; SNAIL; miR-23b	(Bonet et al., 2015)
A0A0B4J2A4	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	Lipid metabolism	-1.737	PPARα; HNF4α	n.a
AA0A024RB23	Diacylglycerol kinase	DGKA	Lipid metabolism	-1.826	PPARγ; Stat5; AP2; Ets1.SP1	n.a
P11216	Glycogen phosphorylase, brain form	PYGB	Starch and sucrose metabolism	-2.018	n.a	(Zhang et al., 2018)
P47989	Xanthine dehydrogenase/oxidase	XDH	Purine metabolism	-2.908	NF-Y	n.a
Q53FA7	Quinone oxidoreductase PIG3	TP53I3	Oxidative stresses and irradiation	-2.911	FOXK2&BAP1	(Alonso et al., 2007; Reka et al., 2014)
Q53X91	Sulfotransferase (Fragment)	SULT1E1	Steroid hormone biosynthesis	-3.010	Nrf2	n.a
ABYXX4	Glutamine synthetase	GLUL	Carbohydrate metabolism	-3.582	ATF4	n.a
<b>Log2(D492M/D492)</b>						
Q53TK1	Cytochrome P450, family 1, subfamily B, polypeptide 1, isoform CRA_a	CYP1B1	Lipid metabolism	3.733	SP1	(Kwon et al., 2016; Shaul et al., 2014)
E7EPM6	Long-chain-fatty-acid-CoA ligase 1	ACSL1	Lipid metabolism	1.712	SP1	(Sánchez-Martínez et al., 2015)
Q6PCE3	Glucose 1,6-bisphosphate synthase	PGM2L1	Carbohydrate metabolism	1.422	ZEB1	n.a
O95394	Phosphoacetylglucosamine mutase	PGM3	Carbohydrate metabolism	1.393	n.a	n.a
Q6FGJ9	Glutathione S-transferase	GSTM3	Glutathione metabolism	1.368	Nrf2	(Zhou et al., 2008)
Q86SF2	N-acetylgalactosaminyltransferase 7	GALNT7	Glycan biosynthesis and metabolism	1.328	miR-30d/30b; miR-214	n.a
Q6FH49	NNMT protein	NNMT	Nicotinate and nicotinamide metabolism	1.181	Stat3	(Eckert et al., 2019; Shaul et al., 2014)
Q8BRR6-2	Isoform 2 of ADP-dependent glucokinase	ADPGK	Glycolysis / Gluconeogenesis	1.173	n.a	(Lee et al., 2016; Song et al., 2018)
O60701	UDP-glucose 6-dehydrogenase	UGDH	Carbohydrate metabolism	1.025	SP1	(Tang et al., 2016; Vergara et al., 2015)
H0UIA1	Acy-CoA synthetase short-chain family member 2, isoform CRA_c	ACSS2	Carbohydrate metabolism	-1.125	SREBF1/2; HIF; TFEB	(Sun et al., 2017)
Q53GQ0	Vary-long-chain 3-oxoacyl-CoA reductase	HSD17B12	Lipid metabolism	-1.142	n.a	n.a
Q53FA7	Quinone oxidoreductase PIG3	TP53I3	Oxidative stresses and irradiation	-1.159	FOXK2&BAP1	(Alonso et al., 2007; Reka et al., 2014)
Q14376	UDP-glucose 4-epimerase	GALE	Galactose metabolism	-1.458	n.a	n.a
Q8LET6	MGST1 protein (Fragment)	MGST1	Glutathione metabolism	-1.522	n.a	(Fischer et al., 2015; Shaul et al., 2014)
P11216	Glycogen phosphorylase, brain form	PYGB	Starch and sucrose metabolism	-1.611	n.a	(Zhang et al., 2018)
P12532	Creatine kinase U-type, mitochondrial	CKMT1A	Arginine and proline metabolism	-1.886	LncRNA n335586&miR-924; EVI1&RUNX1	(Tanaka and Ogishima, 2015)
A0A0B4J2A4	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	Lipid metabolism	-2.038	PPARα; HNF4α	n.a
B4DLR8	NAD(P)H dehydrogenase [quinone] 1	NQO1	Ubiquinone and other terpenoid-quinone biosynthesis	-2.069	Nrf2; NF-Kb	(Fischer et al., 2015; Yang et al., 2017)
Q16831	Uridine phosphorylase 1	UPP1	Pyrimidine metabolism	-2.303	NF-Kb; Oct3/4	(Guan et al., 2019; Wehbe et al., 2012)
AA0A024RB23	Diacylglycerol kinase	DGKA	Lipid metabolism	-2.730	PPARγ; Stat5; AP2; Ets1.SP1	n.a
ABYXX4	Glutamine synthetase	GLUL	Carbohydrate metabolism	-2.741	ATF4	n.a
P47989	Xanthine dehydrogenase/oxidase	XDH	Purine metabolism	-3.159	NF-Y	n.a

## 4.4 Glycan metabolism alters in breast EMT

The enzyme UGDH identified in EMT and GFPT2 identified in EMP are both involved in glycan metabolism. UGDH catalyzes the oxidation of UDP-Glc into UDP-GlcA, while GFPT2 is the rate-limiting enzyme in the HBP to produce UDP-GlcNAc. Glycan, the branched structure that comprises monosaccharide units, is responsible for protein glycosylation, a common PTM. Cellular glycome is diverse and vital for normal cellular functions (Schjoldager et al., 2020). We confirmed the oncogenic properties of UGDH and GFPT2, investigated their influences on the metabolic network, and probed the upstream signaling regulations of these two enzymes.

### 4.4.1 UGDH is highly expressed in mesenchymal cells (*paper I*)

For UGDH, we confirmed the effects of UGDH on patient survival, cell proliferation, cell invasion, and the RNA expression of SNAI1 in the mesenchymal cell lines D492M and D492HER2. We further explored the perturbation of UGDH knockdown on metabolism in mesenchymal cells. Finally, we interrogated the regulatory signaling pathways that control cellular UGDH expression.

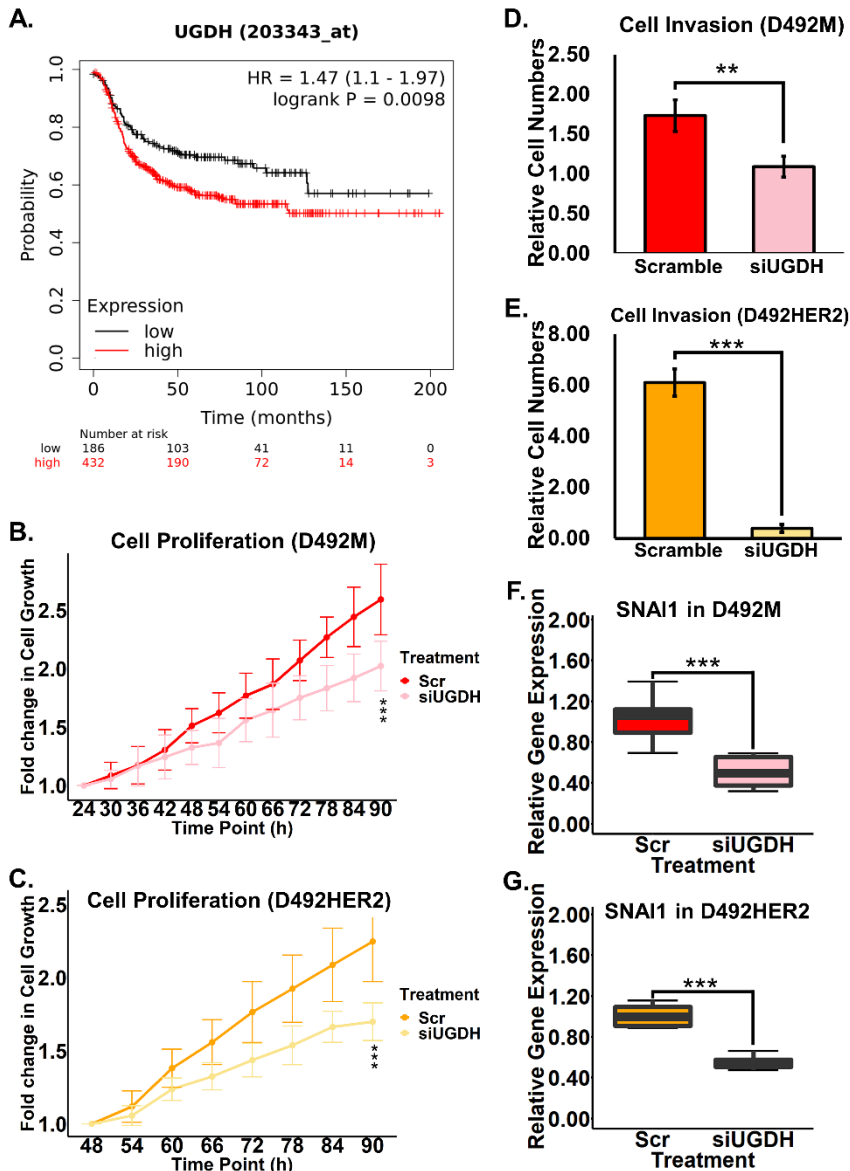
#### ***UGDH knockdown affects patient survival, cell proliferation and invasion, and SNAI1 expression***

UGDH has been reported to affect patient survival (Teoh et al., 2020), cell proliferation (Lin et al., 2020; Oyinlade et al., 2018), cell invasion (Arnold et al., 2019), cell migration (Oyinlade et al., 2018; Teoh et al., 2020), and SNAI1 expression (X. Wang et al., 2019).

The output from the online KM plotter, kmplot.com, revealed that the elevated UGDH expression was negatively correlated with the survival of patients with basal breast cancer, in consonance with the literature (**Figure 18A**). On account of the oncological traits of UGDH, we included two types of breast mesenchymal cells, the non-tumorigenic D492M and tumorigenic D492HER2, to confirm the effects of UGDH on cell proliferation, cell invasion, and SNAI1 expression via siRNA silencing. Same as stated in the literature, knockdown of UGDH impaired cell growth (**Figure 18B-C**) and invasion (**Figure 18D-E**) in both D492M and D492HER2, and the RNA expression of SNAI1 was also hampered by UGDH knockdown (**Figure 18F-G**).

All the mesenchymal cells in the three EMT cell models have increased the expression of UGDH after EMT (**Figure 16**) yet remained non-tumorigenic, which is not directly in accordance with the negative correlation

between UGDH and cancer patient survival (**Figure 18A**) and with the observation that abated UGDH diminished cell proliferation and invasion (**Figure 18B-E**). It suggests that UGDH is not a decisive player in the tumor initiation process, albeit cells with tumorigenicity may depend on UGDH to reinforce malignancy. The inhibition of SNAI1 by decreased UGDH (**Figure 18F-G**) strengthens the connection between UGDH and EMT and pinpoints that UGDH exerts uncanonical regulatory functions, exceeding its catalytic capacity arguably via glycosylation.





**Figure 18.** Functional analysis of the UGDH knockdown in the mesenchymal cells.

**(A)** The expression of UGDH was negatively correlated with the patient survival in basal breast cancer patients based on the Kaplan-Meier plot from kmplot.com. **(B-C)** UGDH knockdown via siRNA decreased cell proliferation in the non-tumorigenic D492M **(B)** and tumorigenic D492HER2 **(C)** cell lines. **(D-E)** UGDH knockdown via siRNA reduced cell invasion in the non-tumorigenic D492M **(D)** and tumorigenic D492HER2 **(E)** cell lines. **(F-G)** The EMT transcription factor SNAI1 was downregulated with the siRNA knockdown of UGDH in the non-tumorigenic D492M **(F)** and tumorigenic D492HER2 **(G)** cell lines. Student's T-test, \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . SNAI1: or Snail, zinc finger protein SNAI1. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

### ***UGDH is associated with the expression of GPC and NAA***

To evaluate the impact of UGDH on metabolism, we treated the mesenchymal cells D492M, HMLEM, and PMC42ET in the three EMT cell models with siRNA that silences UGDH and carried out targeted and untargeted metabolomics analysis. Notwithstanding the knockdown of UGDH, the metabolomes of the cell lines with the same background were clustered together (**Figure 19A-B**), suggesting that the absence of UGDH is unable to achieve distinct metabolic phenotypes. The metabolome of D492M was more comparable to that of HMLEM than PMC42ET (**Figure 19B**), which was in agreement with the proteome clustering (**Figure 7B**). In line with the literature, the substrate of UGDH, UDP-Glc, was increased in all the mesenchymal cell lines with the knockdown of UGDH (**Figure 19C**), while the product of UGDH, UDP-GlcA, was decreased (**Figure 19D**), both of which are crucial for normal cellular functions and impact broadly on cellular activities (Arnold et al., 2019; X. Wang et al., 2019).

We next conducted an untargeted metabolomics analysis to uncover the metabolic influences of UGDH in a systemic manner. The knockdown of UGDH significantly reduced the intracellular level of GPC (**Figure 19E**) and upregulated that of NAA (**Figure 19F**) in all the mesenchymal cell lines, with GPC being affected more prominently. Detections of GPC and NAA in the epithelial D492, non-tumorigenic mesenchymal D492M, and tumorigenic mesenchymal D492HER2 showed that GPC was higher in both D492M and D492HER2 compared with D492 (**Figure 19G**), indicating that GPC may be negatively correlated with UGDH and highly expressed in mesenchymal cells albeit tumorigenicity. The knockdown of UGDH was confirmed to decrease GPC and increase NAA in D492HER2 and MDA-MB-231 (**Figure 20A-D**). To further investigate the connection between GPC and the mesenchymal state, we did a correlation analysis using published datasets in literature (H. Li et al., 2019; Shaul et al., 2014) but did not observe any significant correlation

**(Figure 20E)**, suggesting that the higher GPC levels in the D492 mesenchymal cell lines are irrelevant to their mesenchymal traits. In addition, UGDH knockdown via siRNA did not yield consistent and significant changes to choline and phosphocholine **(Figure 20F-G)**.

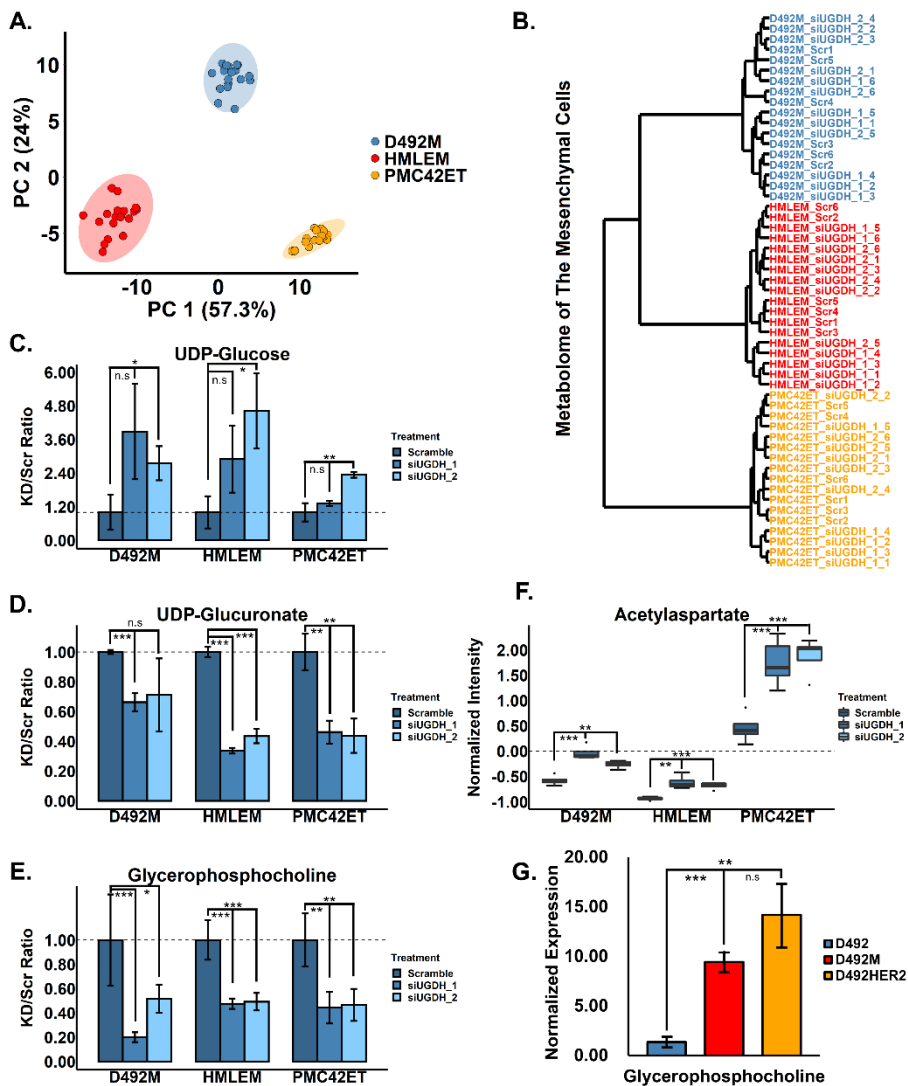
GPC is a choline precursor and a key component in the choline metabolic pathway. Abnormal choline metabolism along with elevated GPC level has emerged as a hallmark of cancer pathogenesis (Sonkar et al., 2019). Increased GPC level has been associated with worse patient survival (M. D. Cao et al., 2012) and is observed in basal-like breast cancer xenograft and ESR-negative breast cancer patients (Giskeødegård et al., 2010; Siver A. Moestue et al., 2010). Abated GPC level with chemotherapies presages better survival in patients with breast cancer (M. D. Cao et al., 2012). GPC is suggested to be involved in EMT, but the mechanism is elusive (Bharti et al., 2018; Koch et al., 2016). Our results have associated GPC with UGDH, which is upregulated in mesenchymal cells **(Figure 19E)**. However, further investigation is needed to clarify the mechanism.

GPC is engaged in choline synthesis, and NAA is intimately related to acetyl-CoA and central carbon metabolism. We conducted *in silico* knockdown of UGDH using tailored GEMs of D492 to query if changes in metabolic fluxes with UGDH knockdown might act on the processing of GPC and NAA (Halldorsson et al., 2017; Karvelsson et al., 2021). Metabolic fluxes of keratan and hyaluronan metabolism, the pentose phosphate pathway, and central carbon metabolic pathways were altered **(Table 8)**, with trivial changes observed to the flux of GPC metabolism.

The affected metabolic pathways predicted by the *in-silico* knockdown of UGDH in GEMs **(Table 8)** suggest a rerouting of glucose flux away from UDP-GlcA formation and into central carbon metabolism, potentially causing the upregulation of NAA **(Figure 19F)**. The negligible impact on GPC implies that decreased GPC with UGDH knockdown is likely caused by changes from altered glycosylation rather than perturbation of the metabolic flux balance. Stable cellular osmotic pressure is critical for cells to carry out normal cell activities for survival. Both GPC and proteoglycans are prominent osmotic regulators functioning as intracellular and extracellular osmolytes, respectively (Okazaki et al., 2019). The knockdown of UGDH causes reduced proteoglycans and extracellular osmotic pressure, which the decreased intracellular GPC may counteract.

UGDH regulates signaling pathways and genes related to lipid metabolism, for example, SNAI1, SIP-1, ERK/MAPK, SIX1, and PPAR $\gamma$

(Arnold et al., 2019; Lin et al., 2020; Teoh et al., 2020; X. Wang et al., 2019). Lipid metabolism, downstream of choline/PtdCho metabolism (E. Iorio et al., 2005), is controlled by the transcription receptor PPAR $\gamma$  (Ahmadian et al., 2013; Tontonoz and Spiegelman, 2008). UGDH inhibits PPAR signaling to regulate lipid metabolism (Arnold et al., 2019). In line with the literature, phosphoproteomics analysis showed higher UGDH expression is associated with downregulated PPAR $\gamma$  signaling (**Figure 20H**), proposing that the knockdown of UGDH decreases GPC via PPAR $\gamma$ . Furthermore, phospholipase A2 group XV (PLA2G15) from the cPLA2 group catalyzes the hydrolysis of phospholipids and possibly the formation of GPC from PtdCho and is regulated by ERK signaling (Menzel et al., 2012; Ulisse et al., 2000). In highly invasive ovarian cancer cells, the knockdown of UGDH diminishes the phosphorylation of ERK (pERK) (Lin et al., 2020). The higher GPC and PLA2G15 in the mesenchymal D492M and D492HER2 (**Figure 19G & Figure 20I**) imply that UGDH affects GPC via pERK-PLA2G15. Collectively, UGDH may indirectly regulate GPC via pERK-PLA2G15 and/or PPAR $\gamma$  to balance the intra- and extracellular osmotic pressure, although further investigation is needed.



**Figure 19.** Metabolomics analysis of the UGDH knockdown in the mesenchymal cells.

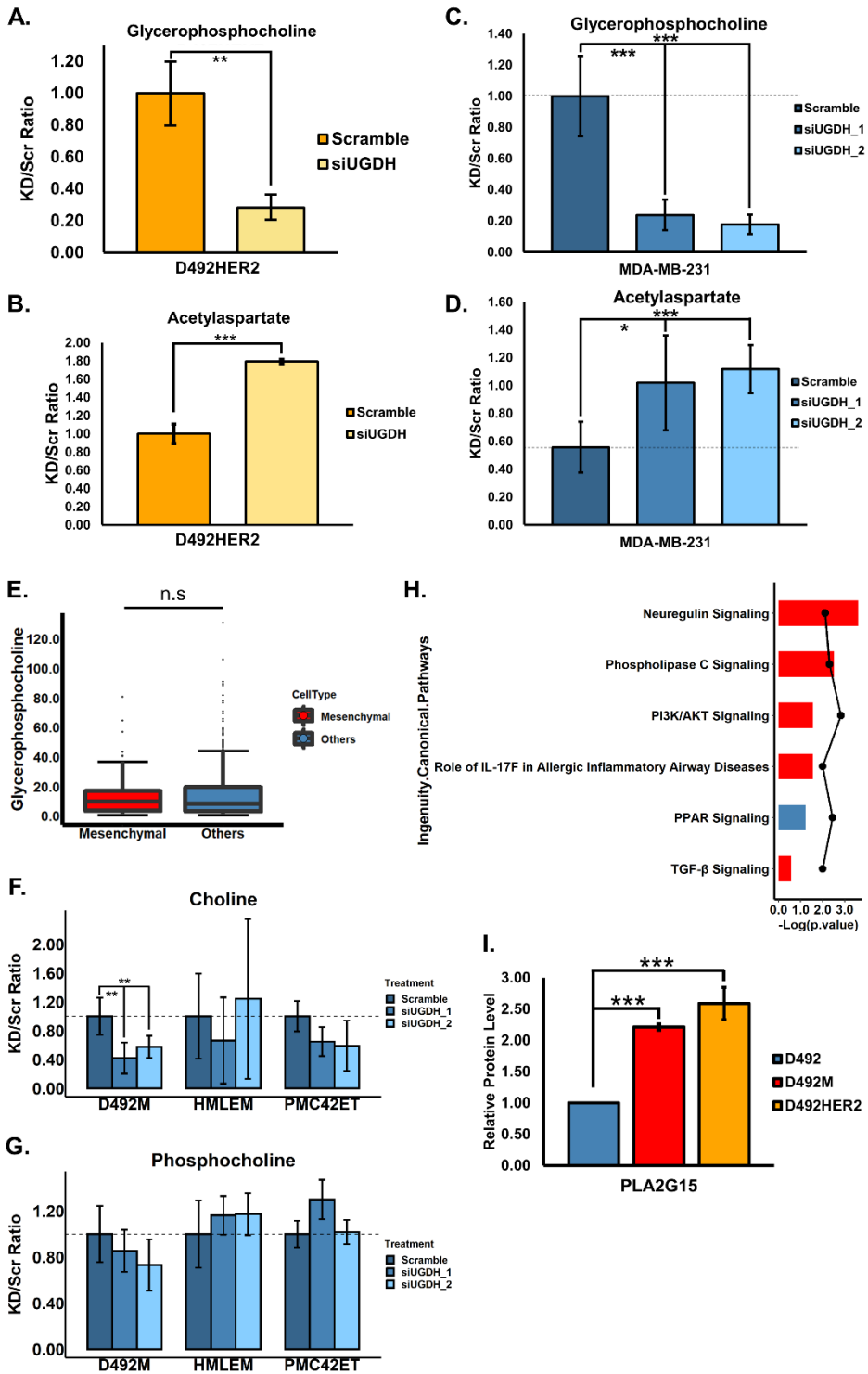
(A) PCA analysis clustered the same cell lines together based on the differences of the metabolomes among different EMT cell models, in spite of UGDH knockdown. (B) The metabolome of the D492 mesenchymal cells was similar to that of the HMLE mesenchymal cells than PMC42. (C) UDP-glucose (UDP-Glc), the substrate of the reaction catalyzed by UGDH, was increased with the siRNA knockdown of UGDH in all mesenchymal cell lines confirmed by two siRNAs; however, one siRNA failed to obtain the significance. (D) UDP-glucuronate (UDP-GlcA), the product of the reaction catalyzed by UGDH, decreased with the siRNA knockdown of UGDH in all the mesenchymal cell lines confirmed by two siRNAs; however, one siRNA in D492M failed to yield significance. (E-F) siRNA knockdown of UGDH significantly decreased the intracellular glycerophosphocholine (GPC) and increased acetylaspertate (NAA) in

all the mesenchymal cell lines confirmed by two siRNAs. **(G)** The intracellular GPC level was higher in the tumorigenic mesenchymal D492HER2 than in the non-tumorigenic D492M and the epithelial D492. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . UGDH: UDP-glucose dehydrogenase. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

**Table 8.** Metabolic pathways affected by UGDH via *in silico* analysis.

*In silico* knockdown of UGDH in GEMs that were built previously on the D492 EMT cell model (Karvelsson et al., 2021) suggested that metabolic pathways with changed metabolic fluxes were involved in keratan, hyaluronan, glycan, central carbon metabolism. UGDH: UDP-glucose dehydrogenase.

Group	Enriched set size	Total set size	p-value	Adjusted p-value
Transport, extracellular	3	859	0	0
Keratan sulfate degradation	71	74	5.25E-101	3.94E-100
Keratan sulfate synthesis	59	59	1.07E-79	5.35E-79
Transport, mitochondrial	1	188	2.26E-07	8.49E-07
Hyaluronan metabolism	5	5	3.38E-06	1.01E-05
Transport, lysosomal	10	33	1.31E-04	3.28E-04
O-glycan synthesis	4	7	0.0012	0.0025
Glyoxylate and dicarboxylate metabolism	3	6	0.0085	0.0159
Pentose phosphate pathway	5	27	0.0454	0.0757
Transport, golgi apparatus	4	25	0.0963	0.1444
Sphingolipid metabolism	2	17	0.2576	0.3512
Citric acid cycle	1	19	0.334	0.3755
Starch and sucrose metabolism	1	8	0.3636	0.3755
Alanine and aspartate metabolism	1	9	0.3755	0.3755
Miscellaneous	1	9	0.3755	0.3755



**Figure 20.** The altered GPC and NAA expression with the siUGDH treatment in D492HER2 and MDA-MB-231 and the possible roles of PPAR signaling and PLA2G15 in regulating GPC.

**(A-D)** In the tumorigenic D492HER2 and malignant MDA-MB-231, the intracellular glycerophosphocholine (GPC) level was decreased **(A, C)**, while the acetylaspartate (NAA) level was increased **(B, D)** after the siUGDH treatment, consistent with the other mesenchymal cell lines. **(E)** Correlation analysis suggested that the GPC level is not correlated with the mesenchymal traits. **(F-G)** Other important metabolites, i.e., choline and phosphocholine, are not consistently and significantly associated with UGDH knockdown. **(H)** Arnold et al. reported that UGDH decreased the PPAR expression to regulate lipid metabolic genes (Arnold et al., 2019). We performed a phosphoproteomics analysis on the D492 EMT cell model to confirm the negative correlation between UGDH and PPAR signaling. The PPAR signaling was downregulated in the mesenchymal D492M cells with highly expressed UGDH. The IPA pathways were listed in a descending order based on the  $-\log_{10}(p \text{ value})$ , and the z-scores of the pathways were plotted by dots/line. Red: overrepresented in the mesenchymal D492M; blue: overrepresented in the epithelial D492. **(I)** PLA2G15 is higher in both D492M and D492HER2 compared with D492 and is an enzyme potentially involved in the hydrolysis of PtdCho into GPC in choline metabolism. cPLA2 is under the regulation of ERK/MAPK, and UGDH has been reported to regulate the phosphorylation of ERK (pERK) (Lin et al., 2020; Menzel et al., 2012; Ulisse et al., 2000), suggesting that the knockdown of UGDH may decrease GPC via pERK and PLA2G15. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . UGDH: UDP-glucose dehydrogenase. PLA2G15: phospholipase A2 group XV. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

### ***UGDH is under the regulation of PDGFRB potentially via NF- $\kappa$ B-p65***

Employing the secretomic data of the D492 EMT model (Steinhaeuser et al., 2020), we interrogated the UGDH signaling regulation. The growth factors IGF, TGF- $\beta$ , and PDGFD were increasingly secreted into the medium by D492M **(Figure 21A)**. We engaged in investigating PDGF signaling in regulating UGDH on account of the highly expressed PDGFRB in the mesenchymal cells **(Figure 21B)** and the secreted PDGFD by D492M **(Figure 21C)** accompanying the dysregulated target of the PDGF signaling PKC **(Figure 21D)**. Both RELA (the gene that codes NF- $\kappa$ B-p65), the downstream target of the PDGFR signaling (Naidu et al., 2017; Shimamura et al., 2002), and UGDH were downregulated with PDGFRB knockdown in the mesenchymal cells **(Figure 21E-G)**. Furthermore, the UGDH RNA level was reduced with RELA knockdown in the mesenchymal cells **(Figure 21H-I)**.

Various regulators are related to UGDH, such as Slit2, SP1, TGF- $\beta$ , hypoxia, p38<sup>MAPK</sup>, LMP2A, and PI3K/Akt, empathizing that UGDH is under the control of a regulatory network (Bontemps et al., 2003; Clarkin et al., 2010; Fernando and Smith, 2020; Pan et al., 2008). The PDGF signaling and its downstream target NF- $\kappa$ B can mediate these regulators and are involved

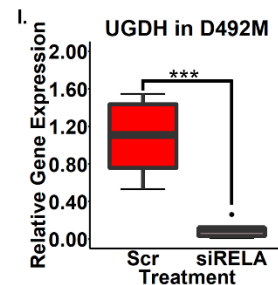
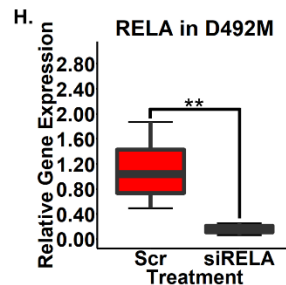
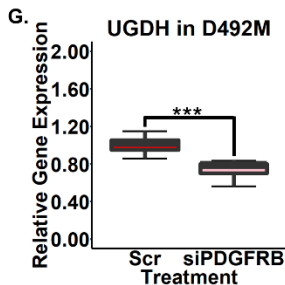
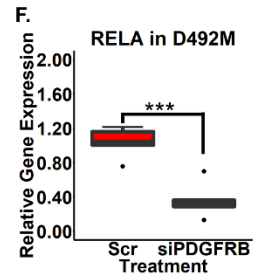
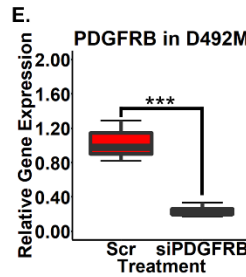
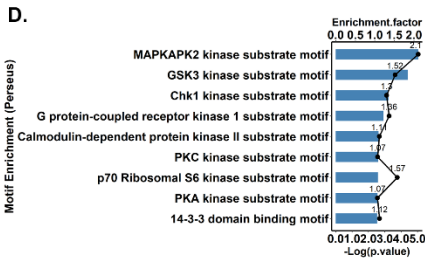
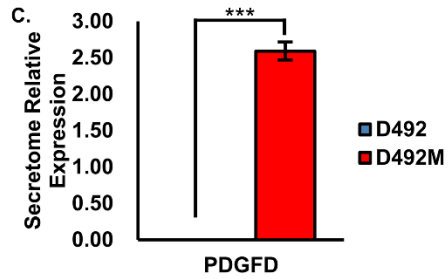
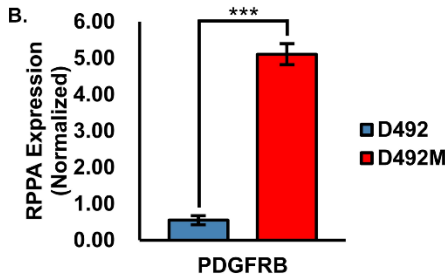
in cancer progression and EMT (Huber et al., 2004; Naidu et al., 2017; Tam et al., 2013). The EGFR-to-PDGFR signaling switch in the formation of CSC and EMT (Tam et al., 2013) is in consonance with the higher level of PDGFRB and secretion of PDGFD in the mesenchymal cells, indicating that the PDGFRB signaling is overrepresented in mesenchymal cells (**Figure 21A-C**). Phospholipase C, PI3K/Akt, and PKC $\alpha$  are common downstream targets of the PDGFR signaling (Tam et al., 2013; H. Wang et al., 2012), and the upregulation of these regulators support the increased PDGFD-PDGFRB signaling in mesenchymal cells (**Figure 20H & Figure 21D**). These findings agree with the literature (Z. Wang et al., 2010; Q. Wu et al., 2013). The inhibitions of PDGFRB and NF- $\kappa$ B-p65 impeded the UGDH expression (**Figure 21E-I**), confirming that PDGFRB regulates UGDH via NF- $\kappa$ B-p65.

Taken together, UGDH contributes to cancer progression and takes part in a collaborative signaling and metabolism network.



**A.**

Protein ID	Protein Name	Gene Name	-Log(pvalue)	FDR	-Log2(D492/D492M)
Q8WX77	Insulin-like growth factor-binding protein-like 1	IGFBPL1	2.633	0.004	-6.055
P22692	Insulin-like growth factor-binding protein 4	IGFBP4	6.401	0.000	-5.008
P61812	Transforming growth factor beta-2	TGFB2	4.302	0.000	-4.524
P01137	Transforming growth factor beta-1	TGFB1	1.773	0.014	-3.392
Q16270	Insulin-like growth factor-binding protein 7	IGFBP7	5.191	0.000	-3.341
Q9GZP0	Platelet-derived growth factor D	PDGFD	4.127	0.000	-3.273
P18065	Insulin-like growth factor-binding protein 2	IGFBP2	2.878	0.003	-1.729
Q15582	Transforming growth factor-beta-induced protein ig-h3	TGFB1	2.606	0.004	-0.967
M0QXF7	Myeloid-derived growth factor (Fragment)	MYDGF	3.098	0.002	-0.848
G3V3X5	Latent-transforming growth factor beta-binding protein 2	LTBP2	2.411	0.005	-0.750
J3KT38	Growth factor receptor-bound protein 2 (Fragment)	GRB2	1.933	0.013	-0.295
H7BYW6	Platelet-derived growth factor subunit A (Fragment)	PDGFA	1.646	0.019	0.505
P08581	Hepatocyte growth factor receptor	MET	1.478	0.024	0.610
P51858	Hepatoma-derived growth factor	HDGF	1.829	0.014	0.647
A0A0C4DH07	Latent-transforming growth factor beta-binding protein 4	LTBP4	2.645	0.004	0.827
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	2.810	0.003	1.728
Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2	EPS8L2	3.919	0.000	1.835
P00533	Epidermal growth factor receptor	EGFR	2.168	0.010	2.485
C9JD84	Latent-transforming growth factor beta-binding protein 1	LTBP1	4.083	0.000	2.501
Q14512	Fibroblast growth factor-binding protein 1	FGFBP1	6.742	0.000	8.075



**Figure 21.** The upstream signaling regulation of UGDH.

(A) The significantly differently secreted growth factors from the cell medium cultured with D492 EMT cells were listed (Permutation-based FDR < 0.05). (B) Based on the RPPA analysis (Barkovskaya et al., 2021), the protein expression of PDGFRB was higher in the mesenchymal cells than the epithelial cells in the D492 model. (C) The growth factor PDGFD was secreted in the mesenchymal cells in the D492 EMT model. (D) The motif enrichment analysis of the phosphorylation sites detected in the phosphoproteomic data yielded plausibly disturbed kinases of the D492 EMT model. PKC kinase activity, among others, was highly enriched in EMT based on the motif enrichment analysis of the phosphorylated proteome in the D492 EMT model. Motif enrichment terms (Enrichment factors  $\geq 2$ ) were reported and ranked based on the  $-\log_{10}(p \text{ value})$  in a descending order. (E) The knockdown efficiency of the PDGFRB siRNA was about 80 % in the D492M cell line. (F) After the siRNA knockdown of PDGFRB in D492M, RELA (the gene that codes NF- $\kappa$ B-p65) was decreased. (G) UGDH was also reduced by the siRNA knockdown of PDGFRB in D492M. (H) The knockdown efficiency of the first RELA siRNA was around 80 % in D492M. (I) UGDH was decreased by the first siRNA knockdown of RELA in D492M. Student's T-test, \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . RPPA: reverse phase protein array; PDGF: platelet-derived growth factor; RELA: or NF- $\kappa$ B, nuclear factor-kappa B; UGDH: UDP-glucose dehydrogenase. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

#### 4.4.2 GFPT2 is associated with cancer malignancy (*paper II*)

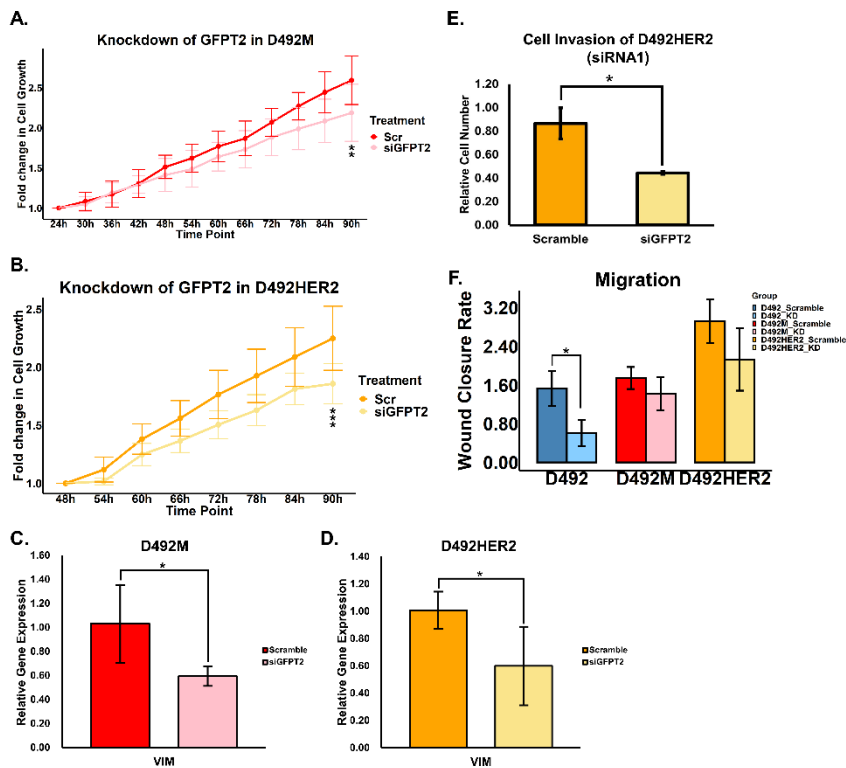
##### ***GFPT2 knockdown affects cell proliferation, EMT, invasion, and migration, and it is highly expressed in claudin-low breast cancer***

Recently, GFPT2 has gained emerging attention in multiple cancer types and has been shown to regulate cell migration and invasion in NSCLC (Szymura et al., 2019), ovarian cancer cell lines (Zhou et al., 2019), and colorectal cancer (L. Liu et al., 2020). GFPT2, as an MMS gene, was shown to decrease the EMT marker vimentin in ovarian cancer (Shaul et al., 2014; Zhou et al., 2019). In light of the reported oncogenic properties of GFPT2 and its influence on EMT, we confirmed that GFPT2 promotes the advancement of breast cancer and affects vimentin in breast mesenchymal cell lines. The proliferation (**Figure 22A-B**) and vimentin expression (**Figure 22C-D**) in the mesenchymal cell lines D492M and D492HER2 were dampened, and the invasiveness of D492HER2 was decreased (**Figure 22E**) by the knockdown of GFPT2. Cell migration was also hindered by GFPT2 knockdown in the three cell lines, although it was unable to secure the significance in the mesenchymal cell lines (**Figure 22F**).

Considering the clinical significance, we further explored the gene expression of GFPT2 across various breast cancer cell lines verified by two data sources and in patients with different breast cancer types. GFPT2 was highly expressed in claudin-low breast cell lines and breast cancer patients compared with the other types (**Figure 23**). This is in compliance with the

previous result from the cell line clustering analysis that the high-GFPT2-expressing D492HER2 clusters with the claudin-low breast cell lines (**Figure 12C**). Claudin-low breast cancer has been deemed a distinct breast cancer subtype as of its discovery. Of late, two publications on *NATURE communications* reported the re-definition and sub-classification of claudin-low breast cancer, questioned the current claudin-low classifier, and concluded that claudin-low breast cancer is heterogeneous and in need of efforts and cautions to define it (Fougner et al., 2020; Pommier et al., 2020). GFPT2 was predicted by Prat et al. to be a claudin-low gene signature in TNBC (Prat et al., 2010), compatible with this study. Additionally, GFPT2 has been reported to promote the malignancy of the KRAS and LKB1 co-mutant NSCLC, which is a type of malignant lung cancer emulating claudin-low breast cancer (H. S. Kim et al., 2013; J. Kim et al., 2020).

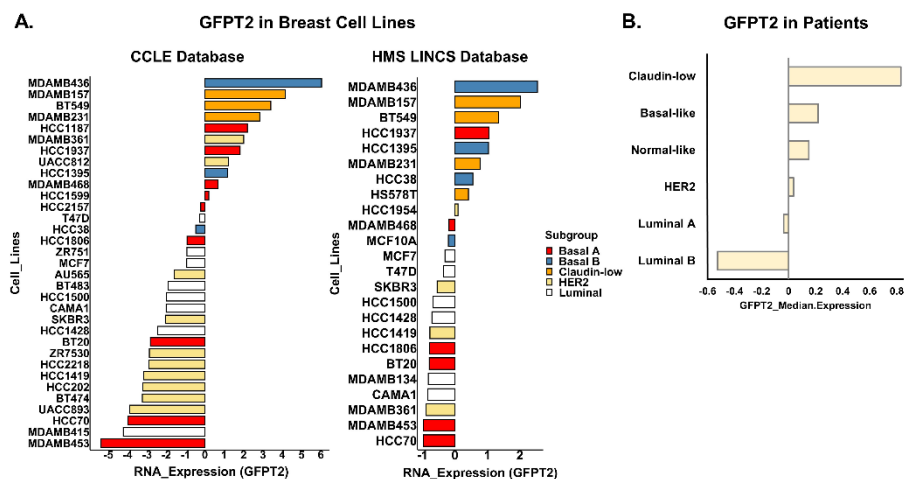
Collectively, these findings indicate that GFPT2 is a marker for claudin-low breast cancer.



**Figure 22.** Functional analysis of the GFPT2 knockdown.

(A-B) The cell proliferation of D492M (A) and D492HER2 (B) were decreased by the siRNA-mediated knockdown of GFPT2 four days from cell seeding. (C-D) The EMT

marker vimentin (VIM) was reduced by the knockdown of GFPT2 in the mesenchymal cell lines D492M (C) and D492HER2 (D). (E) The knockdown of GFPT2 hindered the cell invasion of D492HER2. (F) The knockdown of GFPT2 hampered the cell migration of D492. The decreasing trends of cell growth without significance were seen in D492M and D492HER2. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . KD: Knock Down. GFPT2: glutamine-fructose-6-phosphate transaminase 2. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



**Figure 23.** The GFPT2 expression in different breast cell lines and cancer patients.

Higher expression of GFPT2 was shown in both basal and claudin-low breast cell lines and breast cancer patients. (A) GFPT2 was highly expressed in basal and claudin-low cell lines while lower in HER2-positive and luminal cell lines based on an open-source database – CCLE (left panel) (Ghandi et al., 2019). The trend was confirmed by another open-source database – HMS LINCS (right panel) (Koleti et al., 2017). (B) GFPT2 was highly expressed in patients with claudin-low cancer while lower in patients with HER2-positive and luminal cancer according to the TCGA data (Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)). GFPT2: glutamine-fructose-6-phosphate transaminase 2. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

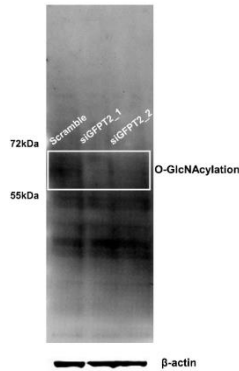
### ***GFPT2 is responsible for tuning the HBP flux***

UDP-GlcNAc is indispensable for the PTM of proteins via O-GlcNAcylation. GFPT2 modulates the availability of UDP-GlcNAc as the precursors for O-GlcNAcylation. The knockdown of GFPT2 dampened the protein O-GlcNAcylation in D492 (Figure 24), supporting the role of GFPT2 in regulating the HBP and the production of UDP-GlcNAc.

We further confirmed the function of GFPT2 in the HBP via metabolomics. We analyzed the metabolomes of the EMP cell model with a targeted metabolomic approach and noticed that the metabolome of D492HER2 was more like that of D492M than D492 (Figure 25A), confirming that D492HER2

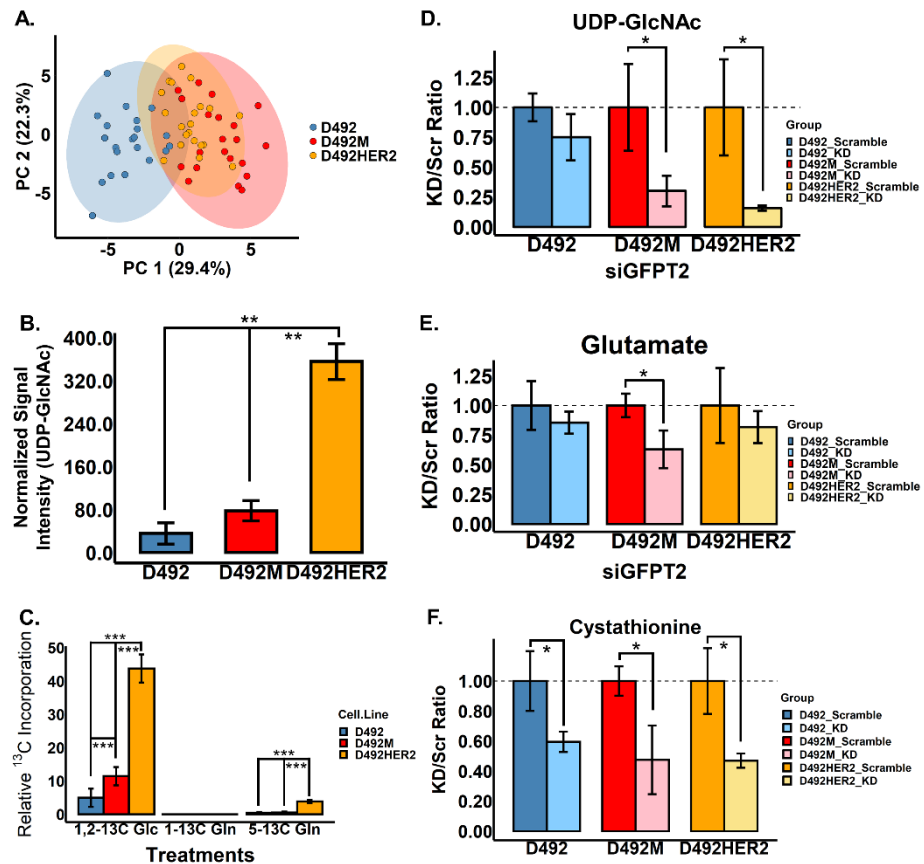
is an intermediate state of D492 and D492M. UPLC-MS analysis of UDP-GlcNAc revealed that D492HER2 expressed a significantly higher amount of UDP-GlcNAc than the other two cell lines, with its intracellular concentration of UDP-GlcNAc increased roughly four-fold than D492M while 10-fold than D492 (**Figure 25B**). The substantial amount of UDP-GlcNAc in D492HER2 is inadequate to deduct the activeness of the metabolic fluxes encircling UDP-GlcNAc. We further traced the enrichment of the  $^{13}\text{C}$  isotopologue labels in UDP-GlcNAc from either 1,2- $^{13}\text{C}$  glucose, 1- $^{13}\text{C}$  glutamine, or 5- $^{13}\text{C}$  glutamine to check the metabolic fluxes into UDP-GlcNAc from different carbon sources. The carbons in UDP-GlcNAc were mainly derived from glucose compared with glutamine (**Figure 25C**), and the relatively small amounts of carbons from 5- $^{13}\text{C}$  glutamine are indicative of an alternate carbon contribution to UDP-GlcNAc through reductive carboxylation. Considering the cellular amount of UDP-GlcNAc, D492HER2 had the highest enrichment of  $^{13}\text{C}$  in UDP-GlcNAc from 1,2- $^{13}\text{C}$  glucose, indicating the absolute metabolite flux of UDP-GlcNAc and the HBP activity are more active in D492HER2. GFPT2 has been reported to be associated with glucose uptake, independent of GLUT1 (Chao et al., 2021; W. Zhang et al., 2018), which may facilitate the increased pool and flux of UDP-GlcNAc in D492HER2. There was no observable  $^{13}\text{C}$  enrichment in UDP-GlcNAc from 1- $^{13}\text{C}$  glutamine on the grounds that the carbon in the first position of glutamine is unlikely to be incorporated into UDP-GlcNAc through either oxidative TCA or citrate-derived cytosolic acetyl-CoA. These data suggest that the higher level and increased metabolic flux of UDP-GlcNAc are in line with the upregulated GFPT2 in D492HER2.

In accordance with the enzymatic functions of GFPT2 that include UDP-GlcNAc production in the HBP and glutamate generation via glutaminolysis (Oki et al., 1999), the knockdown of GFPT2 has led to a significant decrease in intracellular concentration of UDP-GlcNAc in both D492M and D492HER2 (**Figure 25D**) and has dropped in the glutamate level in all the cell lines (**Figure 25E**). It confirms that the upregulated GFPT2 is accountable for the increased HBP flux. Furthermore, we also observed that the treatment of siGFPT2 had significantly diminished the intracellular cystathionine level in all the cell lines (**Figure 25F**).



**Figure 24.** Western blot of protein O-GlcNAcylation with GFPT2 knockdown.

Western blot analysis indicated that the protein O-GlcNAcylation was hindered by the knockdown of GFPT2 in D492 cells treated with two siRNA that interfere with GFPT2, and this experiment used  $\beta$ -actin as the loading control. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



**Figure 25.** Metabolomics analysis of the EMP cell model and the GFPT2 knockdown.

(A) The metabolome of D492HER2 was similar to that of D492M compared with D492. (B) D492HER2 expressed a higher level of UDP-GlcNAc compared with the other cell lines. (C) Carbons incorporation from 1,2-<sup>13</sup>C glucose (Glc) into UDP-GlcNAc were higher compared with 5-<sup>13</sup>C glutamine (Gln) and 1-<sup>13</sup>C Gln in all three cell lines after cell culture for six hours. No carbons were incorporated from 1-<sup>13</sup>C Gln in all three cell lines. D492HER2 had a higher carbon incorporation rate into UDP-GlcNAc from both 1,2-<sup>13</sup>C Glc and 5-<sup>13</sup>C Gln compared with D492 and D492M. The carbon incorporation rate into UDP-GlcNAc from 1,2-<sup>13</sup>C Glc was higher in D492M compared with D492. (D) The production of UDP-GlcNAc in both D492M and D492HER2 was decreased with the knockdown of GFPT2. (E) Glutamate was decreased with the GFPT2 knockdown in the D492 EMT model. (F) Cystathionine was decreased significantly with the knockdown of GFPT2 in all three cell lines. Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . KD: Knock Down. UDP-GlcNAc: UDP-*N*-acetylglucosamine. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

### **GFPT2 is a marker for oxidative stress**

GFPT2 may affect the *de novo* GSH synthesis via glutamine-derived glutamate (Simpson et al., 2012). The knockdown of GFPT2 led to reduced intracellular glutamate levels (**Figure 25E**); however, the impact was insubstantial. We concomitantly observed a reduction in the intracellular cystathionine levels in all the cell lines (**Figure 25F**). Like glutamate, cystathionine is one of the precursors for the *de novo* synthesis of GSH by engaging in the transsulfuration pathway and concurrently producing H<sub>2</sub>S. Knockdown of GFPT2 negligibly affected the GSH level (data not shown), indicating that GFPT2 is not a dominant regulator of GSH.

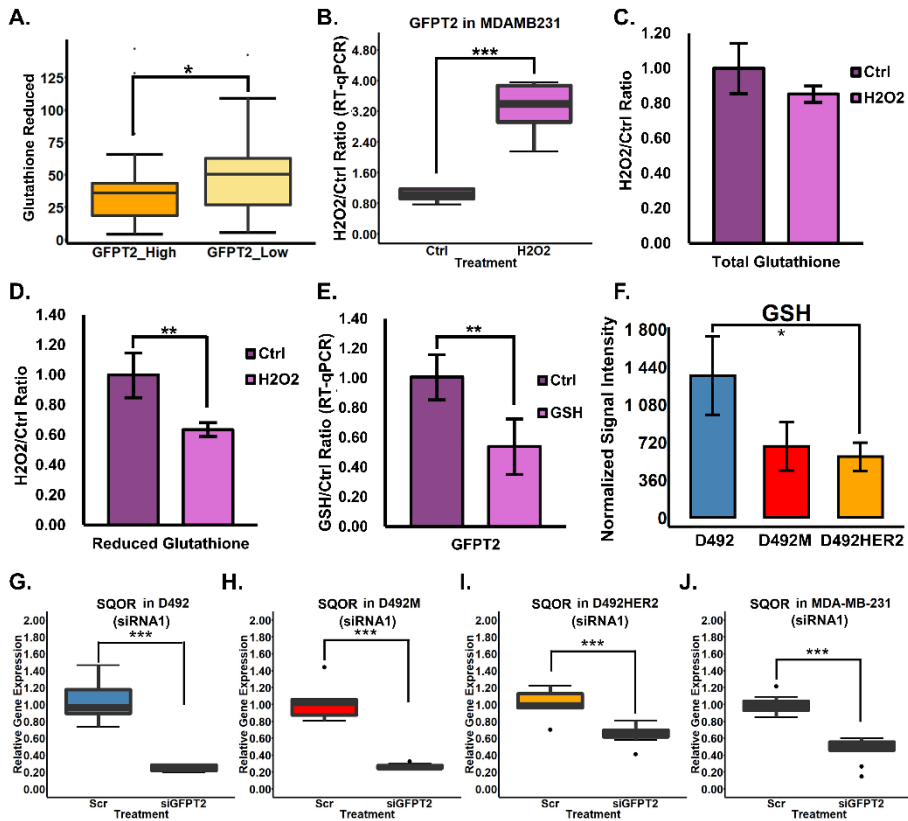
The gene-metabolite correlation analysis using the NCI60 cancer cell line panel implied that GFPT2 and GSH were negatively correlated (**Figure 26A**). There was no correlation observed neither for the oxidized glutathione (GSSG) nor for GFPT1 (data not shown), implying the distinctive relationship between GFPT2 and GSH. Overexpression of GFPT2 increases cell viability and protects cells against the H<sub>2</sub>O<sub>2</sub> treatment (Zitzler et al., 2004). Treatment with H<sub>2</sub>O<sub>2</sub> significantly increased the RNA expression of GFPT2 (**Figure 26B**) and concomitantly reduced the GSH level (**Figure 26D**), with no influences on the total glutathione (**Figure 26C**). In addition, treatment with GSH decreased the RNA expression of GFPT2 (**Figure 26E**). The highest amount of GSH (**Figure 26F**) and lowest expression of GFPT2 in D492 compared with D492M and D492HER2 (**Figure 17**), which are consistent with the increased cell stress in the mesenchymal cells as previously reported (Eiriksson et al., 2018; Halldorsson et al., 2017), support the negative correlation between GFPT2 and GSH. On the grounds of the previous result

that GFPT2 did not regulate GSH, these findings suggest that the RNA expression of GFPT2 is negatively adjusted according to the intracellular level of GSH.

In view of the decreased cystathionine with the knockdown of GFPT2 (**Figure 25F**) and the involvement of cystathionine in the production of H<sub>2</sub>S, GFPT2 may modulate the intracellular H<sub>2</sub>S homeostasis to resist oxidative stress. SQOR locates on the mitochondrial membrane and oxidizes H<sub>2</sub>S and glutathione to regenerate ubiquinol. SQOR was consistently downregulated by the knockdown of GFPT2 (**Figure 26G-J**), indicating that the knockdown of GFPT2 may lead to the diminished production of H<sub>2</sub>S via cystathionine to further hinder the downstream enzyme SQOR. However, further investigation is needed to elucidate the link between GFPT2 and H<sub>2</sub>S. Nonetheless, the decreased cystathionine and SQOR after the knockdown of GFPT2 have associated GFPT2 with H<sub>2</sub>S and mitochondrial metabolism. H<sub>2</sub>S signaling facilitates EMT (Ascencao et al., 2021; M. Wang et al., 2020), the relationship of which is still elusive and controversial (Guo et al., 2016). Mitochondrial dysfunction is involved in EMT (Sessions and Kashatus, 2021) and breast cancer (Lunetti et al., 2019), and GFPT2 is connected to SLP-2, which is involved in mitochondrial regulation (Chao et al., 2021). Dysregulation in GFPT2, H<sub>2</sub>S metabolism, and mitochondrial functions may be part of the partial EMT process in claudin-low breast cancer.

Taken together, these findings suggest that the RNA level of GFPT2 is regulated by redox balance and the increased expression of GFPT2 is a marker for cellular oxidative stress, which plays an important role in EMT (Giannoni et al., 2012) and breast cancer progression (Jezierska-Drutel et al., 2013). GFPT2 may protect cells from oxidative stress via GFPT2-cystathionine-H<sub>2</sub>S-SQOR-mitochondria homeostasis.





**Figure 26.** GFPT2 is a marker for oxidative stress.

(A) A negative correlation between GFPT2 and GSH was observed by the gene-metabolite correlation analysis in the NCI60 cancer cell line panel. (B) The RNA expression of GFPT2 was significantly upregulated in MDA-MB-231 with 2  $\mu$ M H<sub>2</sub>O<sub>2</sub> treatment. (C) The H<sub>2</sub>O<sub>2</sub> treatment did not affect the total glutathione level in MDA-MB-231. (D) The H<sub>2</sub>O<sub>2</sub> treatment significantly decreased the GSH level in MDA-MB-231. (E) The GFPT2 gene expression was significantly downregulated by the treatment of 50 mg/L of GSH in MDA-MB-231. (F) The epithelial D492 had a significantly higher level of GSH than the mesenchymal D492M and D492HER2. (G-J) The knockdown of GFPT2 with siRNA significantly decreased the RNA expression of SQOR in D492 (G), D492M (H), D492HER2 (I), and MDA-MB-231 (J). Student's T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . GSH: reduced glutathione; GFPT2: glutamine-fructose-6-phosphate transaminase 2; SQOR: sulfide quinone oxidoreductase. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

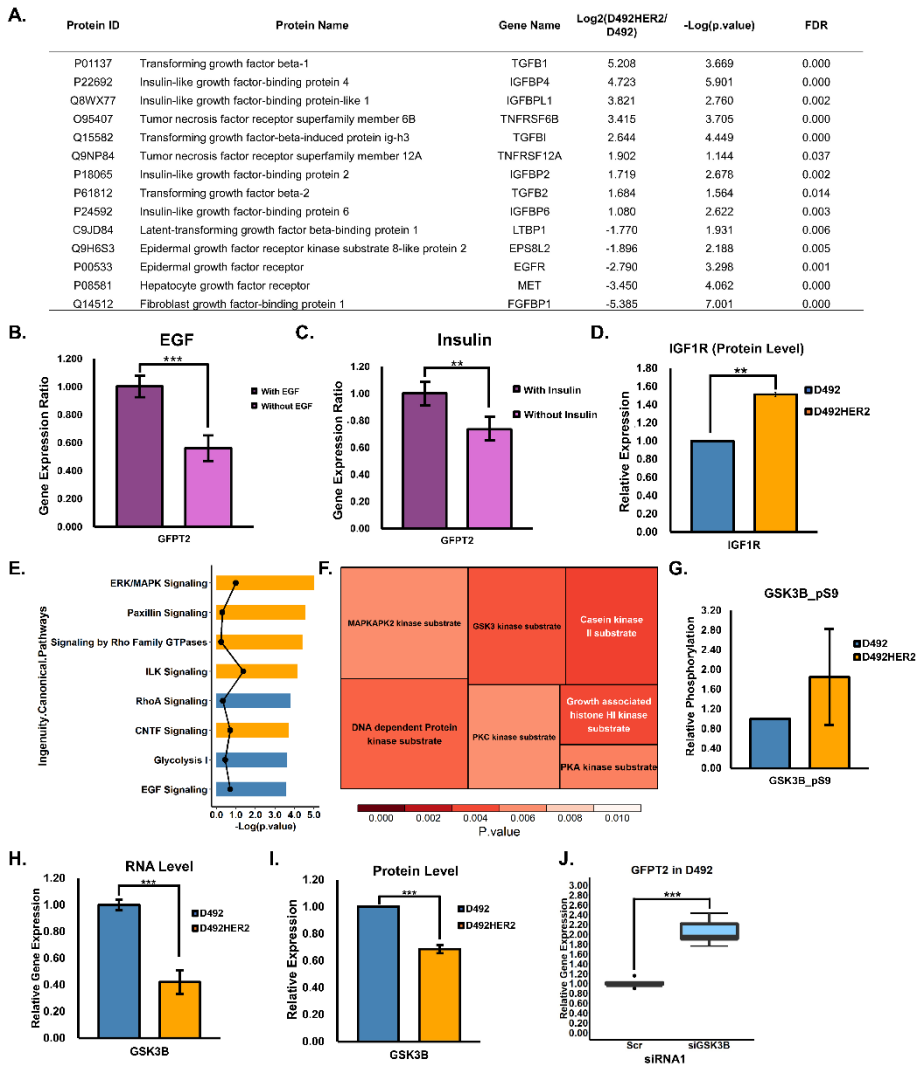
***GFPT2 is under the regulation of EGF and insulin signaling, and it is suppressed by GSK-3 $\beta$*** 

GFPT2 is under the regulation of an intrinsically complex signaling network and related to various growth factors and transcriptional regulators, including EGF, TGF- $\beta$ , TNF, NF- $\kappa$ B, SIRT6, sXBP1, and SLP-2, etc. (Chao et al., 2021; Panarsky et al., 2020; Richani et al., 2014; Szymura et al., 2019; W. Zhang et al., 2018). Mutant KRAS can enhance the flux into HBP via GFPT2 that is intensified by the loss of LKB1 (J. Kim et al., 2020).

The growth factors TGF- $\beta$ , IGF, TNF, and EGF, had secreted differently into the medium between D492HER2 and D492 based on the analysis of the secretomic data (Steinhaeuser et al., 2020) (**Figure 27A**), all of which have been confirmed to regulate GFPT2 apart from IGF (Richani et al., 2014; Szymura et al., 2019; W. Zhang et al., 2018). The removal of insulin and EGF resulted in the decreased expression of GFPT2 (**Figure 27B-C**). The higher expression of IGF1R in D492HER2 (**Figure 27D**) supports that the activities of the IGF signaling are more active in D492HER2.

Phosphoproteomics analysis identified the upregulated ERK/MAPK signaling in D492HER2, a common EGF/IGF downstream pathway. The kinase enrichment analysis also highlighted the enrichment of the GSK-3 $\beta$  and PKC $\alpha$  substrates (**Figure 27E-F**). GSK-3 is one of the central regulators of biological activities and under the inhibition of insulin/IGF1 and ERK/MAPK (Beurel et al., 2015; Cohen and Frame, 2001; Q. Ding et al., 2005; Riis et al., 2020). We further explored GSK-3 $\beta$  inasmuch as it responds to oxidative stress (Niringiyumukiza et al., 2019; Riis et al., 2020; Schafer et al., 2004; C. Yan et al., 2020) and partakes in Wnt signaling, which is of importance to the EMT program (Yook et al., 2006). The phosphorylation of GSK-3 $\beta$  was higher in D492HER2, suggesting GSK-3 $\beta$  is inactivated (**Figure 27G**) (X. M. Xu et al., 2005). GSK-3 $\beta$  were increased on both the RNA and protein levels, indicating D492 has increased activity of the GSK-3 $\beta$  kinase (**Figure 27H-I**). The knockdown of GSK-3 $\beta$  resulted in an increase of the GFPT2 expression (**Figure 27J**), indicating that GSK-3 $\beta$  suppresses GFPT2. These findings suggest that the lower expression of GFPT2 in D492 is mediated by the increased activities of GSK-3 $\beta$ .

In summary, GFPT2 is a tumor promotor highly expressed in mesenchymal cells with partial EMT and a marker for claudin-low breast cancer and oxidative stress. It is under the regulation of insulin and EGF and the kinase GSK-3 $\beta$ .



**Figure 27.** The upstream signaling regulation of GFPT2.

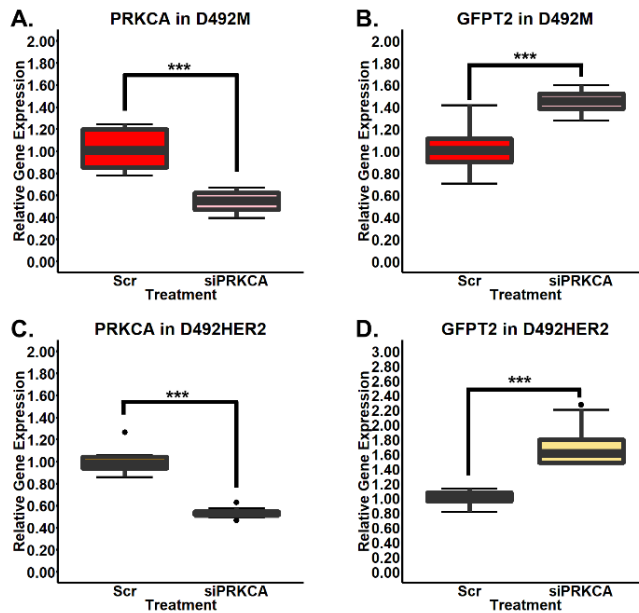
(A) Several growth factors, such as TGF, IGF, TNF, EGF, and FGF, were secreted differently between D492HER2 and D492 based on the analysis of their secretome (FDR < 0.05, Fold change > =2). (B-C) We adapted the MDA-MB-231 cell line in the FBS-free H14 medium to investigate the effects of the growth factors EGF and IGF on GFPT2. The GFPT2 RNA expression was decreased in the MDA-MB-231 cell line with removals of either EGF (B) or insulin (C). (D) The SILAC proteomic data suggested that the protein level of IGF1R was higher in D492HER2 than in D492. (E) The top eight Ingenuity Canonical Pathways from the phosphoproteomics data analysis showed ERK/MAPK signaling was activated in D492HER2. Activated pathways in D492HER2 were labeled in orange, while pathways activated in D492 were marked in blue. Dots were the absolute value of activation Z-scores. Pathways were listed in a descending order based on the -log<sub>10</sub>(p value). (F) Motif enrichment from Perseus (Version 1.6.14.0) suggested that several kinases, such as GSK-3,

Casein kinase, and PKA, were enriched differently in D492HER2 compared with D492. **(G)** According to the phosphoproteomics analysis, GSK-3 $\beta$  was highly phosphorylated at position serine 9, inhibiting the activation of GSK-3 $\beta$  in D492HER2 compared with D492. **(H-I)** Both the RNA **(H)** and protein **(I)** expression of GSK-3 $\beta$  in D492HER2 and D492 showed the higher abundance of GSK-3 $\beta$  in D492 compared with D492HER2. **(J)** The knockdown of GSK-3 $\beta$  with the first siRNA increased the GFPT2 RNA expression in D492. Student's T-test, \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ . EGF: epidermal growth factor; IGF1R: insulin-like growth factor 1 receptor; GSK3B: glycogen synthase kinase 3 beta; GFPT2: glutamine-fructose-6-phosphate transaminase 2. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.

### ***GFPT2 is under the regulation of PKC $\alpha$ (unpublished).***

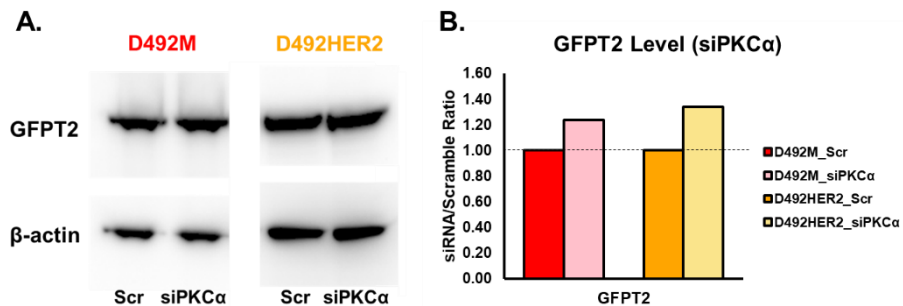
PKC $\alpha$  (gene name: PRKCA) is highly expressed in D492HER2 and D492M than in D492 (**Figure 12A**). PKC $\alpha$  signaling is dysregulated in EMT as stated in literature (Tam et al., 2013), which aligns with the above-discussed data. In light of the central role of PKC $\alpha$  in gene regulation, we explored the potential regulatory functions of PKC $\alpha$  on GFPT2. siRNA-mediated knockdown of PKC $\alpha$  significantly increased the GFPT2 expression on the RNA level in both D492M and D492HER2 (**Figure 28**), which was unexpected since both PKC $\alpha$  and GFPT2 are upregulated in mesenchymal cells, indicating a possible positive association between these two genes. The increase of GFPT2 with PKC $\alpha$  knockdown was confirmed on the protein level in these two cell lines, albeit to a lesser degree (**Figure 29**). The lesser influence of PKC $\alpha$  knockdown on GFPT2 on the protein level than the RNA level is possibly due to the lagged protein expression of GFPT2 following the increased mRNA.

The mechanism by which siRNA-mediated PKC $\alpha$  knockdown increases GFPT2 is unknown. Knockdown of PKC $\alpha$  can protect cells against H<sub>2</sub>O<sub>2</sub> treatment (Saber et al., 2008), which is consistent with the statement that GFPT2 protects cells against H<sub>2</sub>O<sub>2</sub> (Zitzler et al., 2004). This evidence supported that PKC $\alpha$  inhibited GFPT2. Since SIRT6 inhibits GFPT2 (Szymura et al., 2019), and SIRT6 can be degraded by AKT1 phosphorylation (Thirumurthi et al., 2014), One explanation, though purely speculative, is PKC $\alpha$  – AKT1 – SIRT6 – GFPT2. Specifically, PKC $\alpha$  inhibits AKT1 (Hsu et al., 2018) to upregulate SIRT6, leading to GFPT2 downregulation. Further studies are required to confirm this theory and provide more evidence on the mechanism by which PKC $\alpha$  negatively regulates GFPT2.



**Figure 28.** RT-qPCR analysis of GFPT2 with siPKC $\alpha$  treatment.

D492M and D492HER2 were treated with the siRNA targeting PKC $\alpha$ , and the RNA expression of GFPT2 was detected 72 hours after siRNA-mediated PKC $\alpha$  knockdown via RT-qPCR. The knockdown efficiency of the siRNA targeting PKC $\alpha$  was about 50 – 60 % in both D492M (**A**) and D492HER2 (**C**). The RNA expressions of GFPT2 were significantly increased after knockdown of PKC $\alpha$  in both cell lines (**B**, **D**). PRKCA is the gene name of PKC $\alpha$ . Student's T-test, \*\*\*:  $p < 0.001$ . PRKCA: or PKC $\alpha$ , protein kinase C alpha; GFPT2: glutamine-fructose-6-phosphate transaminase 2.



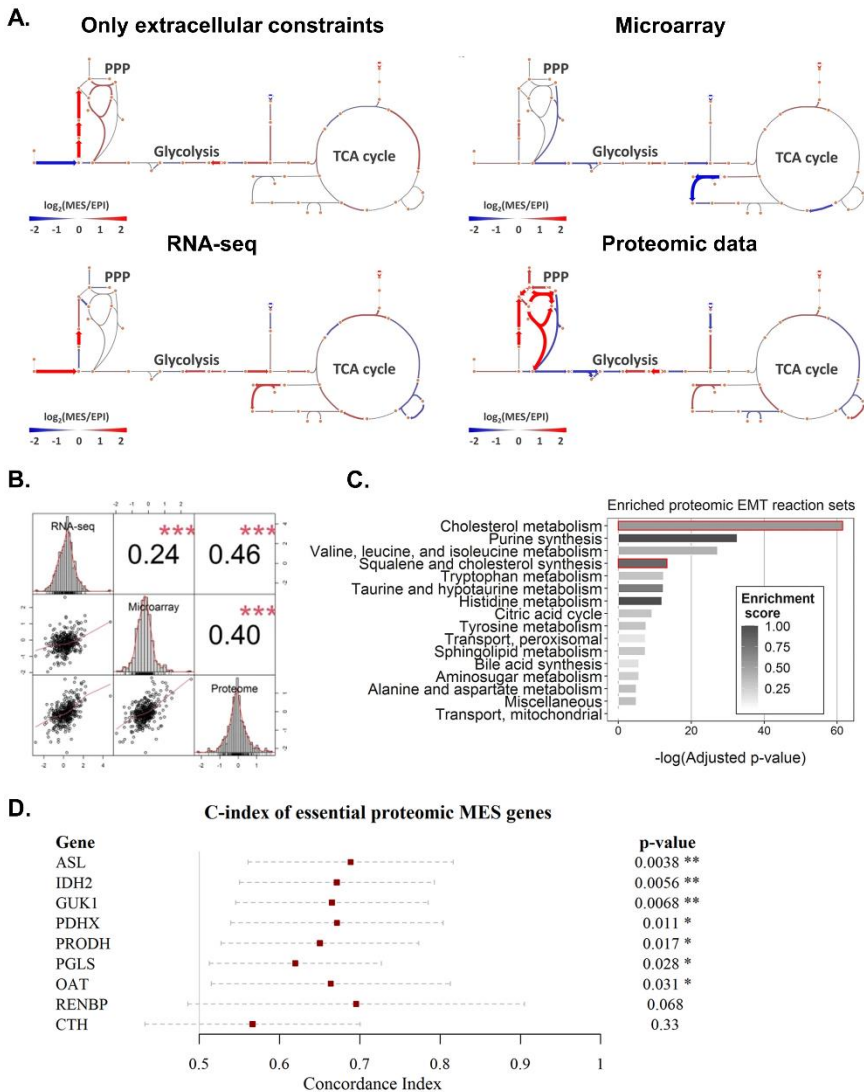
**Figure 29.** Western blot analysis of GFPT2 with siPKC $\alpha$  treatment.

(**A**) D492M and D492HER2 were treated with the siRNA targeting PKC $\alpha$ , and the protein expression of GFPT2 was detected 72 hours after siRNA treatment via western blot. (**B**) Western Blots were quantified by densitometry data in ImageJ software. The protein expressions of GFPT2 were increased after knockdown of PKC $\alpha$  in both cell lines. Loading control:  $\beta$ -actin. PRKCA: or PKC $\alpha$ , protein kinase C alpha; GFPT2: glutamine-fructose-6-phosphate transaminase 2.

## 4.5 Reconstruction and upgrade of the D492 GEMs (*paper III*)

In paper III, the metabolic network of EMT was reconstructed within GEMs using either transcriptomic or proteomic data from the D492 breast EMT cell lines (D492 and D492M cells) based on a previously built model, termed iBreast2886 (**Figure 30**). The proteomic GEM showed higher accuracy in predicting metabolic flux than transcriptomic data. This computational model predicted two features of EMT, which were confirmed in the lab: (1) cholesterol metabolism alteration in EMT; (2) higher dependency on argininosuccinate lyase (ASL) in EMT. iBreast2886 holds potential for interpreting cancer gene expression data in the clinic.

Correlation analysis of the proteomic and RNA-seq data yielded a correlation efficiency of 0.46 (Spearman's rank correlation coefficient), suggesting a low consistency between gene expression on the RNA and protein levels. Two major factors impacting the differences between transcriptome and proteome are PTMs and protein turnover. Metabolic reactions enriched in cholesterol metabolism were identified as key factors in the transition between epithelial and mesenchymal states. Cholesterol synthesis inhibition showed that D492 relies more on this pathway than D492M. Gene essentiality analysis of the proteomic GEM identified ASL as an essential gene in EMT, and application of the breast cancer clinical data in this model revealed that ASL was essential for the survival of patients with ER-negative breast cancer. ASL is a urea cycle enzyme that converts aspartate and citrulline into fumarate and arginine. The roles of ASL in EMT are speculated to be (1) influencing proline synthesis via ornithine aminotransferase (OAT); (2) affecting the fumarate production for the TCA cycle; (3) decreasing the conversion of oxaloacetate (OAA) to aspartate to increase anaplerotic fueling of the TCA cycle.



**Figure 30.** Reconstruction and upgrade of the D492 GEMs using the proteomic dataset.

(A) Different data types generated different metabolic fluxes when they were applied for the reconstruction of the D492 GEMs. Proteomic data-based GEM showed higher accuracy than other data types. (B) Correlation analysis revealed a low consistency of gene expression between the RNA level and the protein level (Spearman's rank correlation coefficient). (C) Metabolic reactions in cholesterol metabolism were highly enriched in the metabolic difference in EMT. (D) Gene essentiality analysis of the proteomic GEM identified ASL, among others, as an essential gene in EMT. \*\*\*:  $p < 0.001$ . ASL: argininosuccinate lyase. Reprinted from (Karvelsson et al., 2021). Reprinted with permission.

## 4.6 Deep mining of the phosphoproteomic data (*unpublished data*)

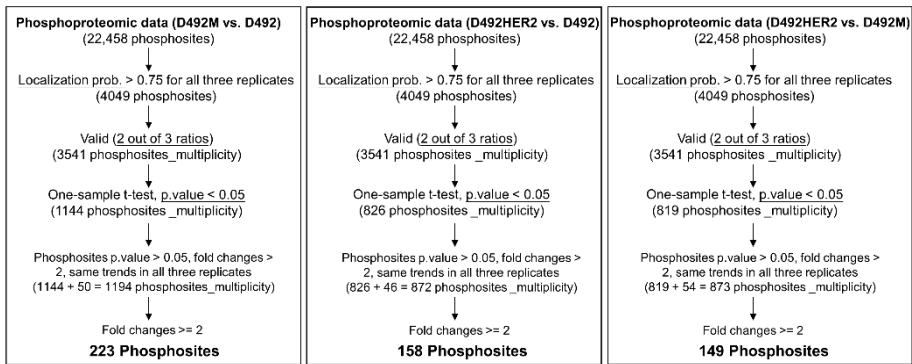
Hereafter, I analyzed and reported the phosphoproteomic data collected from the SILAC experiment for the epithelial D492, mesenchymal D492M, and partial mesenchymal D492HER2, with emphasis on the top dysregulated protein phosphorylation sites between different cell lines and the metabolism-signaling interactions. The phosphoproteomic data were only marginally reported in paper I and II to support the signaling regulation of GFPT2 and UGDH, and an in-depth analysis of this dataset was lacking. On account of the regulatory functions and broad coverage of protein phosphorylation in cellular activities, the following analysis was intended to better understand the phosphorylation alteration in EMT and the different mesenchymal states with a focus on metabolism.

The raw data were filtered to identify the topmost changed phosphosites with valid identification and quantification (**Table 9**). The top altered phosphosites between D492M and D492, D492HER2 and D492, and D492HER2 and D492M were listed in the **Appendix** (223 phosphosites, 158 phosphosites, and 149 phosphosites, respectively).

**Table 9.** Workflow for analysis of the phosphoproteomic dataset.

The phosphoproteomic data were first screened based on the localization probability (localization probability > 0.75 for all three replicates). Two out of three replicates detected with valid SILAC ratios were deemed valid phosphosite identification, and the missing ratios were imputed based on a normal distribution. The three SILAC ratios for each phosphosite were tested in statistical analysis for significance (one-sample Student's T-test p value < 0.05). Phosphosites that did not pass the statistical analysis but with fold changes of more than 2 in all three replicates were also deemed valid differences. Finally, phosphosites were filtered by fold changes (fold change  $\geq$  2). At the end of data filtering and analysis, 223 phosphosites, 158 phosphosites, and 149 phosphosites were significantly different between D492M and D492, D492HER2 and D492, and D492HER2 and D492M, respectively, and these phosphosites were reported in the **Appendix**.





#### 4.6.1 Phosphorylated metabolic alteration in EMT

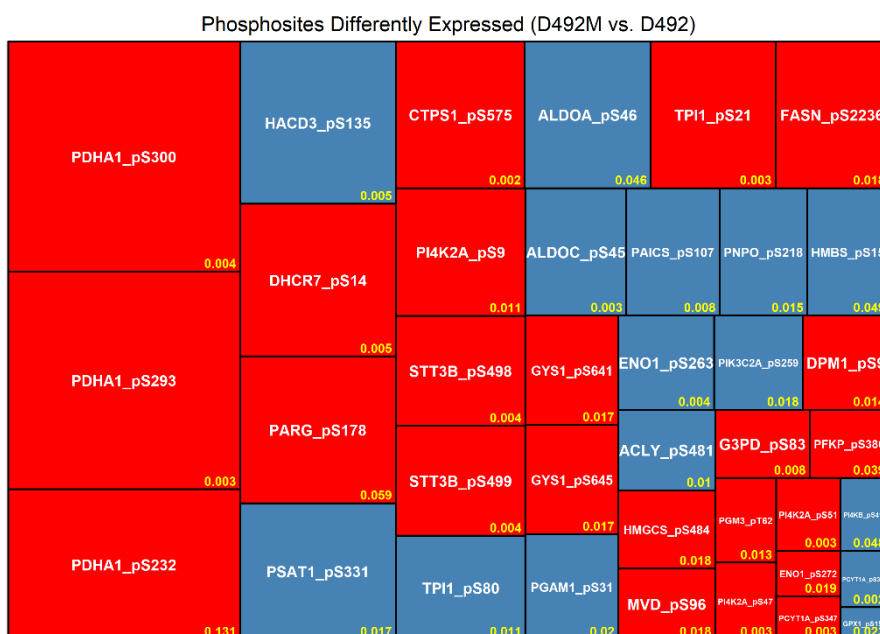
The identified phosphorylated enzymes significantly altered between D492M and D492 were reported (**Figure 31**), and the topmost dysregulated enzymes were plotted (**Figure 32**). Pyruvate dehydrogenase E1 component subunit alpha (PDHA1) was differently phosphorylated at three positions (serine 300, serine 293, and serine 232), which were all higher in D492M compared with D492. In addition, 7-dehydrocholesterol reductase (DHCR7) at serine 14 was also highly phosphorylated in D492M. Conversely, phosphorylations of very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3 (HACD3) at position serine 135 and fructose-bisphosphate aldolase A (ALDOA) at serine 46 were significantly higher in D492.

Under aerobic conditions, the product of glycolysis, pyruvate, is transported into the mitochondrial matrix and completely oxidized to CO<sub>2</sub> in the TCA cycle. During this process, the pyruvate dehydrogenase (PDH) complex, in which PDHA1 plays a key role, catalyzes the irreversible conversion of pyruvate into acetyl-CoA, linking glycolysis and the TCA cycle. The PDH complex is regulated by PDK-1, which is dysregulated in cancer (Flynn et al., 2000). The anaplerosis of pyruvate into the TCA cycle via pyruvate carboxylase (PC) in breast cancer cells can promote cell mobility (Phannasil et al., 2015). The enzymatic activities of PDHA1 can be inhibited by phosphorylation modification of the enzyme at serine 300, 293, and 232 induced by pyruvate dehydrogenase kinase (PDHks or PDKs) (Fujita et al., 2020). Phosphorylation of PDHA1 at serine 293 has been shown to increase cell motility and the expression of EMT markers, such as CDH2, vimentin, and SNAI1, which is in line with the phenotypes of D492M (J. Zhang et al., 2019).

DHCR7 is responsible for the final step in cholesterol production, converting 7-dehydrocholesterol to cholesterol. Both AMPK and protein kinase A (PKA) have been reported to decrease the DHCR7 activities after being treated with kinase inhibitors (Prabhu et al., 2017). Protein Kinase D1 (PRKD1) has been reported to regulate the phosphorylation of DHCR7 at serine 14, which can induce its enzymatic activity (Franz-Wachtel et al., 2012).

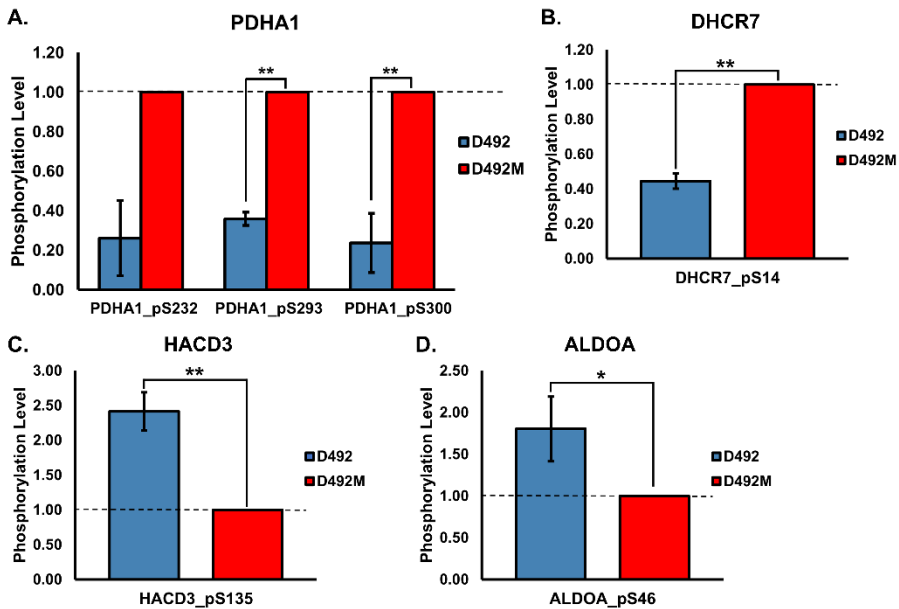
HACD3 is an enzyme in the long-chain fatty acids elongation cycle, and the phosphorylation of HACD3 at serine 135 has been reported before but with no responsible kinases identified (Carrier et al., 2016).

ALDOA participates in glycolysis, where it catalyzes fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate in a reversible manner. The phosphorylation of ALDOA at serine 46 can also be regulated by PRKD1 (Franz-Wachtel et al., 2012).



**Figure 31.** Differently phosphorylated metabolic enzymes between D492M and D492.

The metabolic enzymes differently phosphorylated between D492M and D492 (one-sample Student's T-test p value < 0.05) were plotted. Phosphosites that did not pass the statistical analysis but with fold changes of more than 2 in all three replicates were also deemed valid differences. These metabolic enzymes reported were manually identified. Red: highly expressed in D492M compared with D492. Blue: highly expressed in D492 compared with D492M. Yellow numbers: p values. Rectangle size: the bigger the size, the bigger the difference.



**Figure 32.** Top differences between D492M and D492.

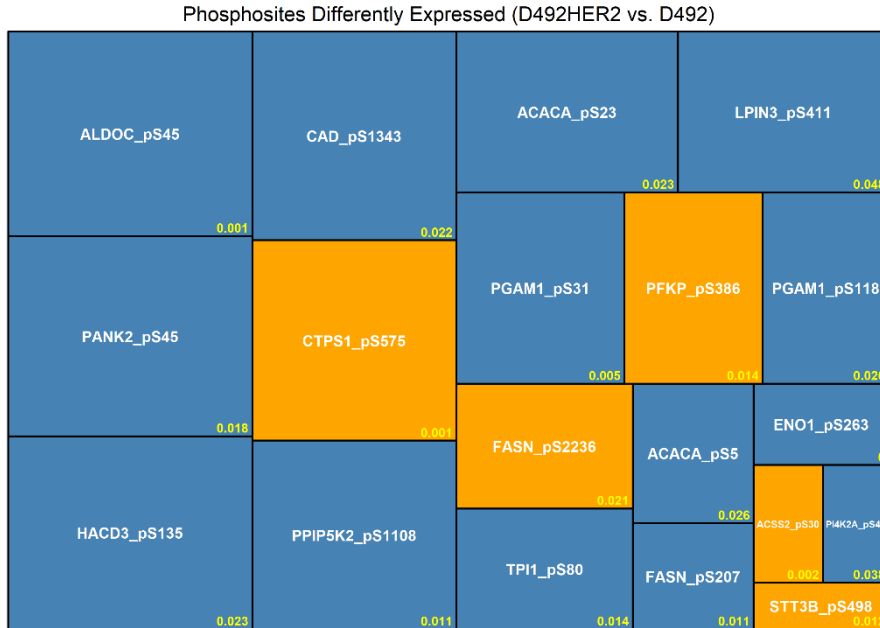
The top differently phosphorylated metabolic enzymes between D492M and D492 were plotted. PDHA1: pyruvate dehydrogenase E1 component subunit alpha; DHCR7: 7-dehydrocholesterol reductase; HACD3: very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3; ALDOA: fructose-bisphosphate aldolase A. One sample T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ .

#### 4.6.2 Phosphorylated metabolic alteration in EMP

The identified phosphorylated enzymes significantly altered between D492HER2 and D492 were reported (**Figure 33**), and the topmost dysregulated enzymes were plotted (**Figure 34**). Fructose-bisphosphate aldolase C (ALDOC) at serine 45, isoform 2 of pantothenate kinase 2 (mitochondrial, PANK2) at serine 45, and HACD3 at serine 135 were highly phosphorylated in D492 compared with D492HER2. On the contrary, the phosphorylation levels of fatty acid synthase (FASN) at serine 2236 and ATP-dependent 6-phosphofructokinase (platelet type, PFKP) at serine 386 were higher in D492HER2.

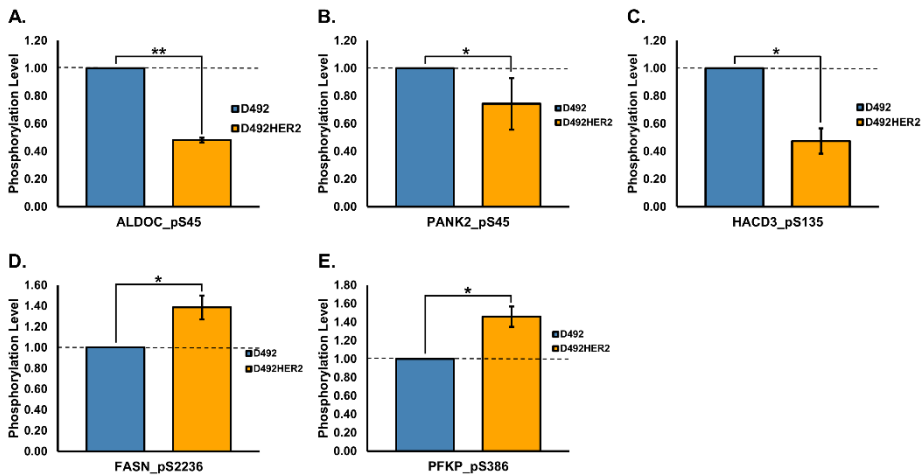
PANK2 plays a critical role in the biosynthesis of coenzyme A. Both HACD3 and fructose-bisphosphate aldolases (ALDOA and ALDOC) were highly phosphorylated in D492 compared with the other two mesenchymal cell lines. FASN is involved in the *de novo* biosynthesis of long-chain saturated fatty acids, while PFKP is responsible for the first committing step of glycolysis, phosphorylating fructose 6-phosphate to fructose 1,6-

bisphosphate. Phosphorylation of PFKP at serine 386 is associated with cell growth and motility, as well as carcinogenic effects (Fan et al., 2021).



**Figure 33.** Differently phosphorylated metabolic enzymes between D492HER2 and D492.

The metabolic enzymes differently phosphorylated between D492HER2 and D492 (one-sample Student's T-test p value < 0.05) were plotted. Phosphosites that did not pass the statistical analysis but with fold changes of more than 2 in all three replicates were also deemed valid differences. These metabolic enzymes reported were manually identified. Orange: highly expressed in D492HER2 compared with D492. Blue: highly expressed in D492 compared with D492HER2. Yellow numbers: p values. Rectangle size: the bigger the size, the bigger the difference.

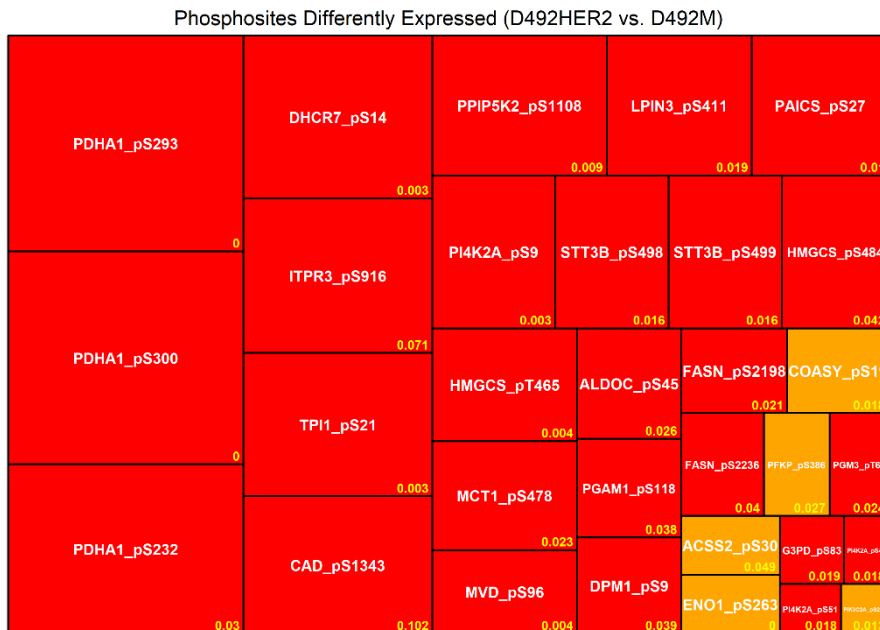


**Figure 34.** Top differences between D492HER2 and D492.

The top differently phosphorylated metabolic enzymes between D492HER2 and D492 were plotted. ALDOC: fructose-bisphosphate aldolase C; PANK2: isoform 2 of pantothenate kinase 2, mitochondrial; HACD3: very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3; FASN: fatty acid synthase; PFKP: ATP-dependent 6-phosphofructokinase, platelet type. One sample T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ .

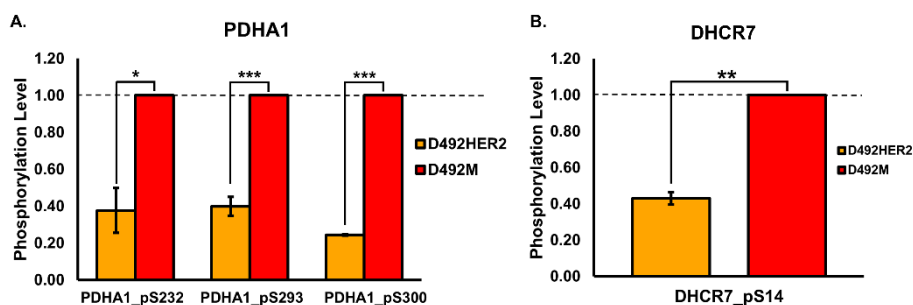
#### 4.6.3 Phosphorylated metabolic alteration in cancer malignancy

The identified phosphorylated enzymes significantly altered between D492HER2 and D492M were reported (**Figure 35**), and the topmost dysregulated enzymes were plotted (**Figure 36**). Similar to the differences between D492M and D492, the phosphorylation levels of PDHA1 and DHCR7 were higher in D492M than in D492HER2.



**Figure 35.** Differently phosphorylated metabolic enzymes between D492HER2 and D492M.

The metabolic enzymes differently phosphorylated between D492HER2 and D492M (one-sample Student's T-test p value < 0.05) were plotted. Phosphosites that did not pass the statistical analysis but with fold changes of more than 2 in all three replicates were also deemed valid differences. These metabolic enzymes reported were manually identified. Red: highly expressed in D492M compared with D492HER2. Orange: highly expressed in D492HER2 compared with D492M. Yellow numbers: p values. Rectangle size: the bigger the size, the bigger the difference.



**Figure 36.** Top differences between D492HER2 and D492M.

The top differently phosphorylated metabolic enzymes between D492HER2 and D492M were plotted. PDHA1: pyruvate dehydrogenase E1 component subunit alpha; DHCR7: 7-dehydrocholesterol reductase. One sample T-test, \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ .

In summary, the topmost dysregulated phosphorylations of enzymes between the epithelial and mesenchymal cell lines are mainly involved in glycolysis, fatty acid synthesis and elongation, cholesterol production, and coenzyme A metabolism. The high phosphorylation levels of PDHA1 suggest that the metabolic fluxes in complete mesenchymal D492M are diverted from the TCA cycle, possibly to facilitate fatty acid metabolism (Halldorsson et al., 2017). The induced activities of DHCR7 in D492M via phosphorylation at serine 14 agree with the need for cholesterol in the mesenchymal cells for cell plasma membrane formation.

#### 4.6.4 Kinases responsible for the identified phosphosites

Proteins are phosphorylated by various kinases. The known kinases responsible for the phosphorylation of metabolic enzymes differentially expressed between two different cell lines were identified based on the platform PhosphoSitePlus: <https://www.phosphosite.org/homeAction.action> (Hornbeck et al., 2015). All metabolic phosphosites reported to be differentially expressed between cell lines were known sites that had been reported in literature except for COASY and HMGCS. However, most of these phosphosites were phosphorylated by unknown kinases. The metabolic phosphosites identified in the previous sections and their responsible kinases are listed in **Table 10**. PDHA1 was highly phosphorylated in D492M compared with both D492 and D492HER2. CAD and PGAM1 were lesser phosphorylated in D492HER2 compared with both D492 and D492M. Pyruvate dehydrogenase kinases (PDHKs) were highly active in the more

complete mesenchymal cell line D492M than both D492 and D492HER2, while Serine/threonine-protein kinase (PAK1) and protein kinase A catalytic subunit  $\alpha$  (PKACA) were less active in the partial mesenchymal D492HER2 compared with D492 and D492M.

**Table 10.** Kinases responsible for known phosphosites.

The kinases responsible for metabolic phosphosites different between two cell lines (reported above) were identified by the database PhosphoSitePlus. The phosphosites with known kinases were reported in this table. PDHA1 can be phosphorylated at serine 300, 293, and 232 by PDHKs and was highly phosphorylated in D492M. The phosphorylation of GYS1 can be induced at position serine 645 by p38 and GSK-3 $\beta$ . CDK1 is the candidate kinase for PIK3C2A phosphorylated at serine 259. CAD can be phosphorylated at position serine 1343 by PKACA, and PGAM1 can be phosphorylated at serine 118 by PAK1. Both CDK1 and CDK2 can phosphorylate PAICS at position serine 27. At last, PKACA is responsible for the phosphorylation of ITPR3 at position serine 916. PDHA1 was highly phosphorylated in D492M compared with both D492 and D492HER2. CAD and PGAM1 were lesser phosphorylated in D492HER2 compared with both D492 and D492M.

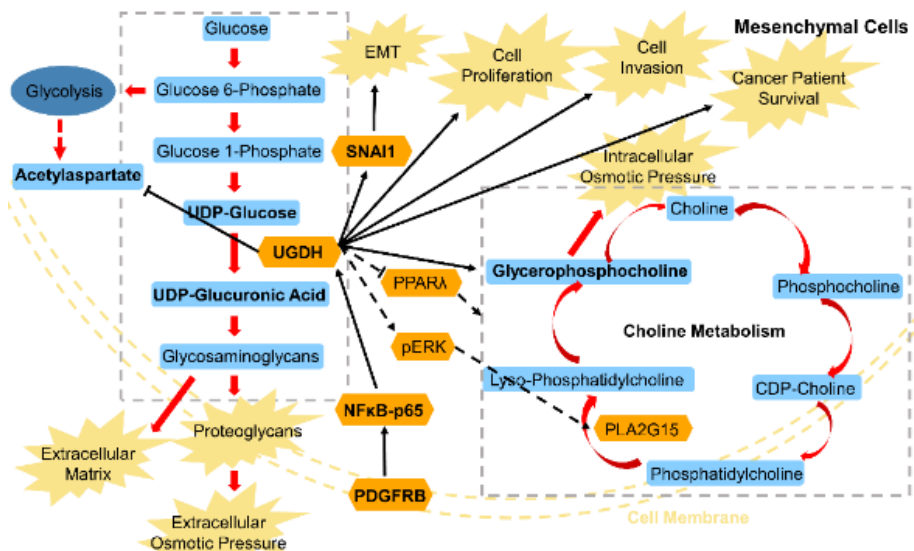
Comparison	Phosphosites	Kinases	Expression
D492M vs. D492	PDHA1_pS300	PDHK4;PDK1;PDHK2;PDHK3;PDHK1	Higer in D492M
	PDHA1_pS293	PDHK4;PDK1;PDHK2;PDHK3;PDHK1	Higer in D492M
	PDHA1_pS232	PDK1;PDHK2;PDHK1	Higer in D492M
	GYS1_pS645	P38B;GSK3B	Higer in D492M
	PIK3C2A_pS259	CDK1	Higer in D492
D492HER2 vs. D492	CAD_pS1343	PKACA	Higher in D492
	PGAM1_pS118	PAK1	Higher in D492
D492HER2 vs. D492M	PDHA1_pS293	PDHK4;PDK1;PDHK2;PDHK3;PDHK1	Higher in D492M
	PDHA1_pS300	PDHK4;PDK1;PDHK2;PDHK3;PDHK1	Higher in D492M
	PDHA1_pS232	PDK1;PDHK2;PDHK1	Higher in D492M
	PAICS_pS27	CDK1;CDK2	Higher in D492M
	PGAM1_pS118	PAK1	Higher in D492M
	ITPR3_pS916	PKACA	Higher in D492M
	CAD_pS1343	PKACA	Higher in D492M



## 5 Summary and conclusions

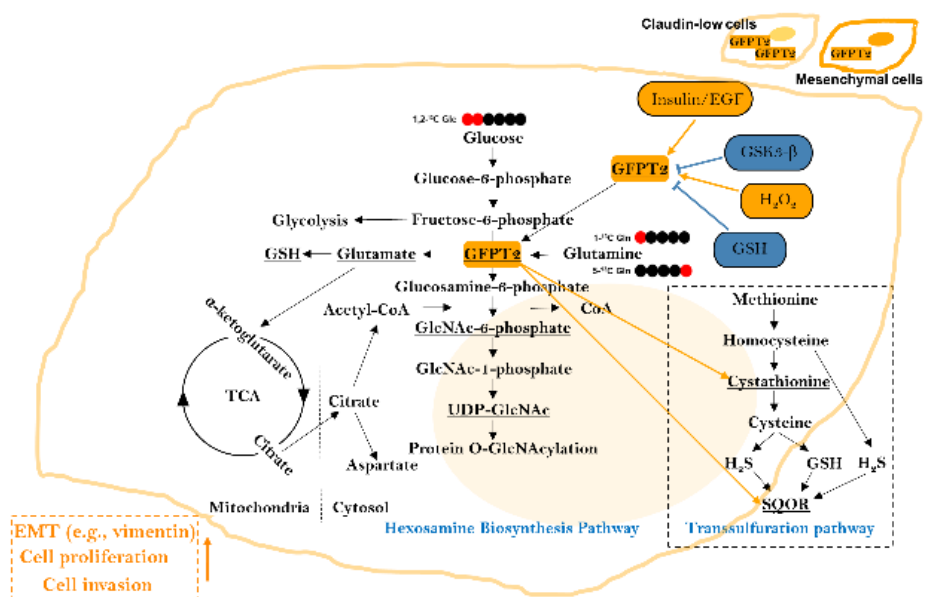
This project first defined the phenotypes of EMT and EMP in breast epithelium metabolically via proteomic and phosphoproteomic approaches, then it identified potential metabolic targets for therapeutic treatments, and finally it investigated the metabolism-signaling relationship in different mesenchymal states. The main findings in this project were summarized in **Figure 37** and **Figure 38**, surrounding UGDH and GFPT2, respectively. Since this project was a discovery-based proteomic study, the main findings, which deserve further attention, could not be followed up within the scope of secured funding and time. The mechanism by which the mesenchymal gene UGDH regulates GPC in choline metabolism is unclear. The explanation for which the partial mesenchymal gene GFPT2 affects cystathionine and SQOR in transsulfuration metabolism is lacking. There were speculation and hypotheses that need the support of experimental data. The signaling regulation of UGDH and GFPT2 proposed in this project needs to be confirmed in other cell types with similar properties and requires to be further elucidated. In addition, the main findings deduced from the phosphoproteomic analysis of the EMP cell models were not followed up.

For future perspective, cell lines with stable knockouts and with overexpression of UGDH and GFPT2 need to be established for further studies of these two genes. Of more interest to us are the roles of UGDH in choline and lipid metabolism and the importance of GFPT2 in cells with defective mitochondria. More generally used breast cell lines and protein assays, for example, western blots, need to be included to confirm the findings from this project and corroborate the conclusions observed on the RNA level. *In vivo* studies are required to verify the cancer promoting roles of these two enzymes especially in tumor metastasis. The identification of PDHK, PAK1, and PKACA from the phosphoproteomic study needs to be followed up by applying similar functional assays, such as siRNA-mediated knockdown, migration, and invasion assays, and more, to evaluate the importance of these kinases in different mesenchymal cell states and explore the potentials of these kinases to distinguish aggressive and non-aggressive mesenchymal cells for breast cancer clinical therapies. The valid findings can be generalized beyond the scope of breast cancer.



**Figure 37.** Summary of the study on spontaneous EMT.

UGDH catalyzes UDP-Glc into UDP-GlcA, an indispensable unit for GAGs, proteoglycans, and ECM. In this study, UGDH was highly expressed in the mesenchymal cells and affected cancer patient survival, mesenchymal cell proliferation and invasion, and the EMT transcription factor SNAI1, and it was under the control of the PDGFRB-NFκB pathway. UGDH knockdown significantly decreased the intracellular GPC level and increased the NAA level in all the mesenchymal cell lines. NAA is closely linked to the central carbon metabolism and is potentially affected by the mass-action effects of UGDH knockdown. GPC is an intracellular osmolyte and part of the choline metabolism. Knockdown of UGDH hindered the formation of proteoglycans and further decreased the extracellular osmotic pressure, which could be counteracted by the reduced intracellular osmotic pressure induced by GPC. We hypothesized that UGDH knockdown affected PPAR $\gamma$ -lipid metabolism and/or pERK-PLA2G15 to regulate GPC and to ease the osmotic stress. UGDH: UDP-glucose dehydrogenase; UDP-Glc: UDP-glucose; UDP-GlcA: UDP-glucuronic acid; GAGs: glycosaminoglycans; ECM: extracellular matrix; EMT: epithelial-mesenchymal transition; SNAI1: zinc finger protein SNAI1; PDGFRB: platelet-derived growth factor receptor B; NFκB: nuclear factor-kappa B; GPC: glycerophosphocholine; NAA: *N*-acetylaspartate; PPAR $\gamma$ : peroxisome proliferator-activated receptor gamma; pERK: phosphorylated extracellular signal-regulated kinase; PLA2G15: phospholipase A2 group XV. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



**Figure 38.** Summary of the study on EMP.

GFPT2 is the rate-limiting enzyme in the HBP and was upregulated in mesenchymal cells, especially in partial EMT represented by the D492HER2 cell line. It affected the EMT program regulator vimentin, the HBP flux, cell growth, and cell invasion, and it was overexpressed in claudin-low breast cancer cell lines and patients with claudin-low breast cancer. GFPT2 was regulated by oxidative stress (H<sub>2</sub>O<sub>2</sub> and GSH) and signaling regulators (insulin and EGF, and GSK-3β). GFPT2 knockdown decreased the intracellular cystathionine level and the SQOR RNA expression in the transsulfuration pathway, indicating that GFPT2 is potentially involved in the H<sub>2</sub>S metabolism and mitochondrial homeostasis. The underscored metabolites were measured in this study. GFPT2: glutamine-fructose-6-phosphate transaminase 2; HBP: hexosamine biosynthetic pathway; EMT: epithelial-mesenchymal transition; H<sub>2</sub>O<sub>2</sub>: hydrogen peroxide; GSH: reduced glutathione; EGF: epidermal growth factor; GSK-3β: glycogen synthase kinase 3 beta; SQOR: sulfide quinone oxidoreductase; H<sub>2</sub>S: hydrogen sulfate. Reprinted from (Q. Wang et al., 2021). Reprinted with permission.



## References

- Aceto, N., Bardia, A., Miyamoto, D. T., Donaldson, M. C., Wittner, B. S., Spencer, J. A., Yu, M., Pely, A., Engstrom, A., Zhu, H., Brannigan, B. W., Kapur, R., Stott, S. L., Shioda, T., Ramaswamy, S., Ting, D. T., Lin, C. P., Toner, M., Haber, D. A., & Maheswaran, S. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*, 158(5), 1110-1122. doi:10.1016/j.cell.2014.07.013
- Ackland, M. L., Michalczyk, A., & Whitehead, R. H. (2001). PMC42, A Novel Model for the Differentiated Human Breast. *Exp Cell Res*, 263(1), 14-22. doi:https://doi.org/10.1006/excr.2000.5106
- Ackland, M. L., Newgreen, D. F., Fridman, M., Waltham, M. C., Arvanitis, A., Minichiello, J., Price, J. T., & Thompson, E. W. (2003). Epidermal growth factor-induced epithelio-mesenchymal transition in human breast carcinoma cells. *Laboratory Investigation*, 83(3), 435-448. doi:10.1097/01.lab.0000059927.97515.fd
- Adam, J., Hatipoglu, E., O'Flaherty, L., Ternette, N., Sahgal, N., Lockstone, H., Baban, D., Nye, E., Stamp, G. W., Wolhuter, K., Stevens, M., Fischer, R., Carmeliet, P., Maxwell, P. H., Pugh, C. W., Frizzell, N., Soga, T., Kessler, B. M., El-Bahrawy, M., Ratcliffe, P. J., & Pollard, P. J. (2011). Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell*, 20(4), 524-537. doi:10.1016/j.ccr.2011.09.006
- Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347-355. doi:10.1038/nature19949
- Aharonov, A., Shakked, A., Umansky, K. B., Savidor, A., Genzelinakh, A., Kain, D., Lendengolts, D., Revach, O. Y., Morikawa, Y., Dong, J., Levin, Y., Geiger, B., Martin, J. F., & Tzahor, E. (2020). ERBB2 drives YAP activation and EMT-like processes during cardiac regeneration. *Nat Cell Biol*, 22(11), 1346-1356. doi:10.1038/s41556-020-00588-4
- Ahmadian, M., Suh, J. M., Hah, N., Liddle, C., Atkins, A. R., Downes, M., & Evans, R. M. (2013). PPARgamma signaling and metabolism: the good, the bad and the future. *Nat Med*, 19(5), 557-566. doi:10.1038/nm.3159

- Aigner, K., Descovich, L., Mikula, M., Sultan, A., Dampier, B., Bonné, S., van Roy, F., Mikulits, W., Schreiber, M., Brabletz, T., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., & Eger, A. (2007). The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression. *FEBS Lett*, *581*(8), 1617-1624. doi:10.1016/j.febslet.2007.03.026
- Akbani, R., Ng, P. K., Werner, H. M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J. Y., Yoshihara, K., Li, J., Ling, S., Seviour, E. G., Ram, P. T., Minna, J. D., Diao, L., Tong, P., Heymach, J. V., Hill, S. M., Dondelinger, F., Stadler, N., Byers, L. A., Meric-Bernstam, F., Weinstein, J. N., Broom, B. M., Verhaak, R. G., Liang, H., Mukherjee, S., Lu, Y., & Mills, G. B. (2014). A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun*, *5*, 3887. doi:10.1038/ncomms4887
- Akella, N. M., Ciraku, L., & Reginato, M. J. (2019). Fueling the fire: emerging role of the hexosamine biosynthetic pathway in cancer. *BMC Biol*, *17*(1), 52. doi:10.1186/s12915-019-0671-3
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, *100*(7), 3983. doi:10.1073/pnas.0530291100
- Al-Mukh, H., Baudoin, L., Bouaboud, A., Sanchez-Salgado, J. L., Maraqa, N., Khair, M., Pagesy, P., Bismuth, G., Niedergang, F., & Issad, T. (2020). Lipopolysaccharide Induces GFAT2 Expression to Promote O-Linked beta-N-Acetylglucosaminylation and Attenuate Inflammation in Macrophages. *J Immunol*, *205*(9), 2499-2510. doi:10.4049/jimmunol.2000345
- Alfaro, J. A., Bohlander, P., Dai, M., Filius, M., Howard, C. J., van Kooten, X. F., Ohayon, S., Pomorski, A., Schmid, S., Aksimentiev, A., Anslyn, E. V., Bedran, G., Cao, C., Chinappi, M., Coyaud, E., Dekker, C., Dittmar, G., Drachman, N., Eelkema, R., Goodlett, D., Hentz, S., Kalathiya, U., Kelleher, N. L., Kelly, R. T., Kelman, Z., Kim, S. H., Kuster, B., Rodriguez-Larrea, D., Lindsay, S., Maglia, G., Marcotte, E. M., Marino, J. P., Masselon, C., Mayer, M., Samaras, P., Sarthak, K., Sepiashvili, L., Stein, D., Wanunu, M., Wilhelm, M., Yin, P., Meller, A., & Joo, C. (2021). The emerging landscape of single-molecule protein sequencing technologies. *Nat Methods*, *18*(6), 604-617. doi:10.1038/s41592-021-01143-1

- Anderson, K. A., Madsen, A. S., Olsen, C. A., & Hirschev, M. D. (2017). Metabolic control by sirtuins and other enzymes that sense NAD<sup>+</sup>, NADH, or their ratio. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1858(12), 991-998. doi:https://doi.org/10.1016/j.bbabi.2017.09.005
- Arnold, J. M., Gu, F., Ambati, C. R., Rasaily, U., Ramirez-Pena, E., Joseph, R., Manikkam, M., San Martin, R., Charles, C., Pan, Y., Chatterjee, S. S., Den Hollander, P., Nagi, C., Sikora, A. G., Rowley, D., Putluri, N., Karanam, B., Mani, S. A., & Sreekumar, A. (2019). UDP-glucose 6-dehydrogenase regulates hyaluronic acid production and promotes breast cancer progression. *Oncogene*. doi:10.1038/s41388-019-0885-4
- Ascencao, K., Dilek, N., Augsburger, F., Panagaki, T., Zuhra, K., & Szabo, C. (2021). Pharmacological induction of mesenchymal-epithelial transition via inhibition of H2S biosynthesis and consequent suppression of ACLY activity in colon cancer cells. *Pharmacol Res*, 165, 105393. doi:10.1016/j.phrs.2020.105393
- Askari, M., Kordi-Tamandani, D. M., Almadani, N., McElreavey, K., & Totonchi, M. (2019). Identification of a homozygous GFPT2 variant in a family with asthenozoospermia. *Gene*, 699, 16-23. doi:10.1016/j.gene.2019.02.060
- Bacigalupo, M. L., Manzi, M., Espelt, M. V., Gentilini, L. D., Compagno, D., Laderach, D. J., Wolfenstein-Todel, C., Rabinovich, G. A., & Troncoso, M. F. (2015). Galectin-1 triggers epithelial-mesenchymal transition in human hepatocellular carcinoma cells. *J Cell Physiol*, 230(6), 1298-1309. doi:10.1002/jcp.24865
- Bae, S. N., Arand, G., Azzam, H., Pavasant, P., Torri, J., Frandsen, T. L., & Thompson, E. W. (1993). Molecular and cellular analysis of basement membrane invasion by human breast cancer cells in Matrigel-based in vitro assays. *Breast Cancer Res Treat*, 24(3), 241-255. doi:10.1007/bf01833264
- Bakir, B., Chiarella, A. M., Pitarresi, J. R., & Rustgi, A. K. (2020). EMT, MET, Plasticity, and Tumor Metastasis. *Trends Cell Biol*, 30(10), 764-776. doi:10.1016/j.tcb.2020.07.003
- Barkovskaya, A., Seip, K., Hilmarsdottir, B., Maelandsmo, G. M., Moestue, S. A., & Itkonen, H. M. (2019). O-GlcNAc Transferase Inhibition Differentially Affects Breast Cancer Subtypes. *Sci Rep*, 9(1), 5670. doi:10.1038/s41598-019-42153-6
- Barkovskaya, A., Goodwin, C. M., Seip, K., Hilmarsdottir, B., Pettersen, S., Stalneck, C., Engebraaten, O., Briem, E., Der,

- C. J., Moestue, S. A., Gudjonsson, T., Maelandsmo, G. M., & Prasmickaite, L. (2021). Detection of phenotype-specific therapeutic vulnerabilities in breast cells using a CRISPR loss-of-function screen. *Mol Oncol*. doi:10.1002/1878-0261.12951
- Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., & García De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol*, 2(2), 84-89. doi:10.1038/35000034
- Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nat Med*, 23(10), 1124-1134. doi:10.1038/nm.4409
- Beauchef, G., Kypriotou, M., Chadjichristos, C., Widom, R. L., Poree, B., Renard, E., Moslemi, S., Wegrowski, Y., Maquart, F. X., Pujol, J. P., & Galera, P. (2005). c-Krox down-regulates the expression of UDP-glucose dehydrogenase in chondrocytes. *Biochem Biophys Res Commun*, 333(4), 1123-1131. doi:10.1016/j.bbrc.2005.06.020
- Bedi, U., Mishra, V. K., Wasilewski, D., Scheel, C., & Johnsen, S. A. (2014). Epigenetic plasticity: a central regulator of epithelial-to-mesenchymal transition in cancer. *Oncotarget*, 5(8), 2016-2029. doi:10.18632/oncotarget.1875
- Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., Schafer, R., van Diest, P., Voest, E., van Oudenaarden, A., Vrisekoop, N., & van Rheenen, J. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep*, 14(10), 2281-2288. doi:10.1016/j.celrep.2016.02.034
- Belke, D. D. (2011). Swim-exercised mice show a decreased level of protein O-GlcNAcylation and expression of O-GlcNAc transferase in heart. *J Appl Physiol (1985)*, 111(1), 157-162. doi:10.1152/jappphysiol.00147.2011
- Bergers, G., & Fendt, S. M. (2021). The metabolism of cancer cells during metastasis. *Nat Rev Cancer*, 21(3), 162-180. doi:10.1038/s41568-020-00320-2
- Bertucci, F., Ng, C. K. Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J. C., Dien, A. T., Adnani, Y., Kamal, M., Garnier, S., Meurice, G., Jimenez, M., Dogan, S., Verret, B., Chaffanet, M., Bachelot, T., Campone, M., Lefeuvre, C., Bonnefoi, H., Dalenc, F., Jacquet, A., De Filippo, M. R., Babbar, N., Birnbaum, D., Filleron, T., Le Tourneau, C., & André, F. (2019). Genomic characterization of metastatic breast



- cancers. *Nature*, 569(7757), 560-564. doi:10.1038/s41586-019-1056-z
- Beurel, E., Grieco, S. F., & Jope, R. S. (2015). Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacol Ther*, 148, 114-131. doi:10.1016/j.pharmthera.2014.11.016
- Bharti, S. K., Mironchik, Y., Wildes, F., Penet, M. F., Goggins, E., Krishnamachary, B., & Bhujwalla, Z. M. (2018). Metabolic consequences of HIF silencing in a triple negative human breast cancer xenograft. *Oncotarget*, 9(20), 15326-15339. doi:10.18632/oncotarget.24569
- Bhat-Nakshatri, P., Appaiah, H., Ballas, C., Pick-Franke, P., Goulet, R., Jr., Badve, S., Srour, E. F., & Nakshatri, H. (2010). SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype. *BMC Cancer*, 10, 411. doi:10.1186/1471-2407-10-411
- Bhatia, S., Monkman, J., Blick, T., Duijf, P. H., Nagaraj, S. H., & Thompson, E. W. (2019). Multi-Omics Characterization of the Spontaneous Mesenchymal-Epithelial Transition in the PMC42 Breast Cancer Cell Lines. *Journal of clinical medicine*, 8(8), 1253. doi:10.3390/jcm8081253
- Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., & Gianni, L. (2016). Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*, 13(11), 674-690. doi:10.1038/nrclinonc.2016.66
- Bianchini, G., De Angelis, C., Licata, L., & Gianni, L. (2022). Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat Rev Clin Oncol*, 19(2), 91-113. doi:10.1038/s41571-021-00565-2
- Biarc, J., Gonzalo, P., Mikaelian, I., Fattet, L., Deygas, M., Gillet, G., Lemoine, J., & Rimokh, R. (2014). Combination of a discovery LC-MS/MS analysis and a label-free quantification for the characterization of an epithelial-mesenchymal transition signature. *J Proteomics*, 110, 183-194. doi:10.1016/j.jprot.2014.05.026
- Bogner-Strauss, J. G. (2017). N-Acetylaspartate Metabolism Outside the Brain: Lipogenesis, Histone Acetylation, and Cancer. *Frontiers in endocrinology*, 8(240). doi:10.3389/fendo.2017.00240
- Bontemps, Y., Vuillermoz, B., Antonicelli, F., Perreau, C., Danan, J. L., Maquart, F. X., & Wegrowski, Y. (2003). Specific protein-1 is a universal regulator of UDP-glucose dehydrogenase expression: its positive involvement in transforming growth

- factor-beta signaling and inhibition in hypoxia. *J Biol Chem*, 278(24), 21566-21575. doi:10.1074/jbc.M209366200
- Brabletz, T., Kalluri, R., Nieto, M. A., & Weinberg, R. A. (2018). EMT in cancer. *Nat Rev Cancer*, 18(2), 128-134. doi:10.1038/nrc.2017.118
- Brandes, A. H., Ward, C. S., & Ronen, S. M. (2010). 17-allylamino-17-demethoxygeldanamycin treatment results in a magnetic resonance spectroscopy-detectable elevation in choline-containing metabolites associated with increased expression of choline transporter SLC44A1 and phospholipase A2. *Breast Cancer Res*, 12(5), R84. doi:10.1186/bcr2729
- Briem, E., Ingthorsson, S., Traustadottir, G. A., Hilmarsdottir, B., & Gudjonsson, T. (2019). Application of the D492 Cell Lines to Explore Breast Morphogenesis, EMT and Cancer Progression in 3D Culture. *Journal of Mammary Gland Biology and Neoplasia*, 10.1007/s10911-018-09424-w. doi:10.1007/s10911-018-09424-w
- Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet*, 21(10), 630-644. doi:10.1038/s41576-020-0258-4
- Bullen, J. W., Balsbaugh, J. L., Chanda, D., Shabanowitz, J., Hunt, D. F., Neumann, D., & Hart, G. W. (2014). Cross-talk between two essential nutrient-sensitive enzymes: O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *J Biol Chem*, 289(15), 10592-10606. doi:10.1074/jbc.M113.523068
- Butti, R., Das, S., Gunasekaran, V. P., Yadav, A. S., Kumar, D., & Kundu, G. C. (2018). Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges. *Mol Cancer*, 17(1), 34. doi:10.1186/s12943-018-0797-x
- Cailleau, R., Olivé, M., & Cruciger, Q. V. (1978). Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro*, 14(11), 911-915. doi:10.1007/bf02616120
- Camarda, R., Zhou, A. Y., Kohnz, R. A., Balakrishnan, S., Mahieu, C., Anderton, B., Eyob, H., Kajimura, S., Tward, A., Krings, G., Nomura, D. K., & Goga, A. (2016). Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nature Medicine*, 22(4), 427-432. doi:10.1038/nm.4055

- Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., & Nieto, M. A. (2000). The transcription factor Snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology*, 2(2), 76-83. doi:10.1038/35000025
- Cao, M. D., Sitter, B., Bathen, T. F., Bofin, A., Lønning, P. E., Lundgren, S., & Gribbestad, I. S. (2012). Predicting long-term survival and treatment response in breast cancer patients receiving neoadjuvant chemotherapy by MR metabolic profiling. *NMR Biomed*, 25(2), 369-378. doi:10.1002/nbm.1762
- Cao, M. D., Cheng, M., Rizwan, A., Jiang, L., Krishnamachary, B., Bhujwala, Z. M., Bathen, T. F., & Glunde, K. (2016). Targeting choline phospholipid metabolism: GDPD5 and GDPD6 silencing decrease breast cancer cell proliferation, migration, and invasion. *NMR Biomed*, 29(8), 1098-1107. doi:10.1002/nbm.3573
- Cao, W., Li, J., Hao, Q., Vadgama, J. V., & Wu, Y. (2019). AMP-activated protein kinase: a potential therapeutic target for triple-negative breast cancer. *Breast Cancer Research*, 21(1), 29. doi:10.1186/s13058-019-1107-2
- Carrier, M., Joint, M., Lutting, R., Page, A., & Rochette-Egly, C. (2016). Phosphoproteome and Transcriptome of RA-Responsive and RA-Resistant Breast Cancer Cell Lines. *PLoS One*, 11(6), e0157290. doi:10.1371/journal.pone.0157290
- Chao, D., Ariake, K., Sato, S., Ohtsuka, H., Takadate, T., Ishida, M., Masuda, K., Maeda, S., Miura, T., Mitachi, K., Yu, X. J., Fujishima, F., Mizuma, M., Nakagawa, K., Morikawa, T., Kamei, T., & Unno, M. (2021). Stomatin-like protein 2 induces metastasis by regulating the expression of a rate-limiting enzyme of the hexosamine biosynthetic pathway in pancreatic cancer. *Oncol Rep*, 45(6). doi:10.3892/or.2021.8041
- Chen, B., Brown, K. A., Lin, Z., & Ge, Y. (2018). Top-Down Proteomics: Ready for Prime Time? *Anal Chem*, 90(1), 110-127. doi:10.1021/acs.analchem.7b04747
- Chen, C., Zhao, S., Karnad, A., & Freeman, J. W. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. *Journal of Hematology & Oncology*, 11(1), 64. doi:10.1186/s13045-018-0605-5
- Chen, D., Dang, B. L., Huang, J. Z., Chen, M., Wu, D., Xu, M. L., Li, R., & Yan, G. R. (2015). MiR-373 drives the epithelial-to-mesenchymal transition and metastasis via the miR-373-TXNIP-HIF1 $\alpha$ -TWIST signaling axis in breast cancer.

- Chen, L., Gibbons, D. L., Goswami, S., Cortez, M. A., Ahn, Y. H., Byers, L. A., Zhang, X., Yi, X., Dwyer, D., Lin, W., Diao, L., Wang, J., Roybal, J., Patel, M., Ungewiss, C., Peng, D., Antonia, S., Mediavilla-Varela, M., Robertson, G., Suraokar, M., Welsh, J. W., Erez, B., Wistuba, II, Chen, L., Peng, D., Wang, S., Ullrich, S. E., Heymach, J. V., Kurie, J. M., & Qin, F. X. (2014). Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun*, 5, 5241. doi:10.1038/ncomms6241
- Cheng, M., Rizwan, A., Jiang, L., Bhujwala, Z. M., & Glunde, K. (2017). Molecular Effects of Doxorubicin on Choline Metabolism in Breast Cancer. *Neoplasia*, 19(8), 617-627. doi:10.1016/j.neo.2017.05.004
- Cheung, K. J., Gabrielson, E., Werb, Z., & Ewald, A. J. (2013). Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell*, 155(7), 1639-1651. doi:10.1016/j.cell.2013.11.029
- Chiaradonna, F., Ricciardiello, F., & Palorini, R. (2018). The Nutrient-Sensing Hexosamine Biosynthetic Pathway as the Hub of Cancer Metabolic Rewiring. *Cells*, 7(6). doi:10.3390/cells7060053
- Christen, S., Lorendeau, D., Schmieder, R., Broekaert, D., Metzger, K., Veys, K., Elia, I., Buescher, J. M., Orth, M. F., Davidson, S. M., Grünewald, T. G., De Bock, K., & Fendt, S. M. (2016). Breast Cancer-Derived Lung Metastases Show Increased Pyruvate Carboxylase-Dependent Anaplerosis. *Cell Rep*, 17(3), 837-848. doi:10.1016/j.celrep.2016.09.042
- Clarkin, C. E., Allen S Fau - Kuiper, N. J., Kuiper Nj Fau - Wheeler, B. T., Wheeler Bt Fau - Wheeler-Jones, C. P., Wheeler-Jones Cp Fau - Pitsillides, A. A., & Pitsillides, A. A. (2010). Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and release, control sulfated GAG synthesis, and promote chondrogenesis. *J Cell Physiol*(1097-4652 (Electronic)).
- Clarkin, C. E., Allen, S., Kuiper, N. J., Wheeler, B. T., Wheeler-Jones, C. P., & Pitsillides, A. A. (2011). Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and release, control sulfated GAG synthesis, and promote chondrogenesis. *J Cell Physiol*, 226(3), 749-761. doi:10.1002/jcp.22393

- Clarkin, C. E., Allen, S., Wheeler-Jones, C. P., Bastow, E. R., & Pitsillides, A. A. (2011). Reduced chondrogenic matrix accumulation by 4-methylumbelliferone reveals the potential for selective targeting of UDP-glucose dehydrogenase. *Matrix Biol*, *30*(3), 163-168. doi:10.1016/j.matbio.2011.01.002
- Clevers, H. (2011). The cancer stem cell: premises, promises and challenges. *Nat Med*, *17*(3), 313-319. doi:10.1038/nm.2304
- Cohen, P., & Frame, S. (2001). The renaissance of GSK3. *Nature Reviews Molecular Cell Biology*, *2*(10), 769-776. doi:10.1038/35096075
- Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., & van Roy, F. (2001). The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell*, *7*(6), 1267-1278. doi:10.1016/s1097-2765(01)00260-x
- Coomer, M., & Essop, M. F. (2014). Differential hexosamine biosynthetic pathway gene expression with type 2 diabetes. *Mol Genet Metab Rep*, *1*, 158-169. doi:10.1016/j.ymgmr.2014.03.003
- Counihan, J. L., Grossman, E. A., & Nomura, D. K. (2018). Cancer Metabolism: Current Understanding and Therapies. *Chem Rev*, *118*(14), 6893-6923. doi:10.1021/acs.chemrev.7b00775
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367-1372. doi:10.1038/nbt.1511
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*, *10*(4), 1794-1805. doi:10.1021/pr101065j
- Cox, J., Hein, M. Y., Lubner, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, *13*(9), 2513-2526. doi:10.1074/mcp.M113.031591
- Cox, J., Hein, M. Y., Lubner, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, *13*(9), 2513-2526. doi:10.1074/mcp.M113.031591

- de Heer, E. C., Jalving, M., & Harris, A. L. (2020). HIFs, angiogenesis, and metabolism: elusive enemies in breast cancer. *J Clin Invest*, *130*(10), 5074-5087. doi:10.1172/JCI137552
- De Livera, A. M., Olshansky, G., Simpson, J. A., & Creek, D. J. (2018). NormalizeMets: assessing, selecting and implementing statistical methods for normalizing metabolomics data. *Metabolomics*, *14*(5), 54. doi:10.1007/s11306-018-1347-7
- Deen, A. J., Arasu, U. T., Pasonen-Seppanen, S., Hassinen, A., Takabe, P., Wojciechowski, S., Karna, R., Rilla, K., Kellokumpu, S., Tammi, R., Tammi, M., & Oikari, S. (2016). UDP-sugar substrates of HAS3 regulate its O-GlcNAcylation, intracellular traffic, extracellular shedding and correlate with melanoma progression. *Cell Mol Life Sci*, *73*(16), 3183-3204. doi:10.1007/s00018-016-2158-5
- Denkert, C., Liedtke, C., Tutt, A., & von Minckwitz, G. (2017). Molecular alterations in triple-negative breast cancer—the road to new treatment strategies. *The Lancet*, *389*(10087), 2430-2442. doi:10.1016/s0140-6736(16)32454-0
- Ding, Q., Xia, W., Liu, J. C., Yang, J. Y., Lee, D. F., Xia, J., Bartholomeusz, G., Li, Y., Pan, Y., Li, Z., Bargou, R. C., Qin, J., Lai, C. C., Tsai, F. J., Tsai, C. H., & Hung, M. C. (2005). Erk associates with and primes GSK-3beta for its inactivation resulting in upregulation of beta-catenin. *Mol Cell*, *19*(2), 159-170. doi:10.1016/j.molcel.2005.06.009
- Ding, X., Liu, H., Yuan, Y., Zhong, Q., & Zhong, X. (2022). Roles of GFPT2 Expression Levels on the Prognosis and Tumor Microenvironment of Colon Cancer. *Frontiers in Oncology*, *12*. doi:10.3389/fonc.2022.811559
- Dongre, A., & Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*, *20*(2), 69-84. doi:10.1038/s41580-018-0080-4
- Duan, J. J., Cai, J., Guo, Y. F., Bian, X. W., & Yu, S. C. (2016). ALDH1A3, a metabolic target for cancer diagnosis and therapy. *Int J Cancer*, *139*(5), 965-975. doi:10.1002/ijc.30091
- Duband, J. L., Dufour, S., Hatta, K., Takeichi, M., Edelman, G. M., & Thiery, J. P. (1987). Adhesion molecules during somitogenesis in the avian embryo. *J Cell Biol*, *104*(5), 1361-1374. doi:10.1083/jcb.104.5.1361
- Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A.,

- Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., Goodacre, R., & The Human Serum Metabolome, C. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7), 1060-1083. doi:10.1038/nprot.2011.335
- Dupuy, F., Tabariès, S., Andrzejewski, S., Dong, Z., Blagih, J., Annis, M. G., Omeroglu, A., Gao, D., Leung, S., Amir, E., Clemons, M., Aguilar-Mahecha, A., Basik, M., Vincent, E. E., St-Pierre, J., Jones, R. G., & Siegel, P. M. (2015). PDK1-Dependent Metabolic Reprogramming Dictates Metastatic Potential in Breast Cancer. *Cell Metab*, 22(4), 577-589. doi:10.1016/j.cmet.2015.08.007
- Durbin, B. P., Hardin, J. S., Hawkins, D. M., & Rocke, D. M. (2002). A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, 18(suppl\_1), S105-S110. doi:10.1093/bioinformatics/18.suppl\_1.S105
- Eades, G., Yao, Y., Yang, M., Zhang, Y., Chumsri, S., & Zhou, Q. (2011). miR-200a regulates SIRT1 expression and epithelial to mesenchymal transition (EMT)-like transformation in mammary epithelial cells. *J Biol Chem*, 286(29), 25992-26002. doi:10.1074/jbc.M111.229401
- Edelman, G. M., Gallin, W. J., Delouvé, A., Cunningham, B. A., & Thiery, J. P. (1983). Early epochal maps of two different cell adhesion molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 80(14), 4384-4388. doi:10.1073/pnas.80.14.4384
- Eiriksson, F. F., Rolfsson, O., Ogmundsdottir, H. M., Haraldsson, G. G., Thorsteinsdottir, M., & Halldorsson, S. (2018). Altered plasmalogen content and fatty acid saturation following epithelial to mesenchymal transition in breast epithelial cell lines. *The International Journal of Biochemistry & Cell Biology*, 103, 99-104. doi:10.1016/j.biocel.2018.08.003
- Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L., Popescu, N. C., Hahn, W. C., & Weinberg, R. A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev*, 15(1), 50-65. doi:10.1101/gad.828901
- Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., Verfaillie, C., Grunewald, T. G. P., & Fendt, S. M. (2017). Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat Commun*, 8, 15267. doi:10.1038/ncomms15267

- Elia, I., Doglioni, G., & Fendt, S. M. (2018). Metabolic Hallmarks of Metastasis Formation. *Trends Cell Biol*, 28(8), 673-684. doi:10.1016/j.tcb.2018.04.002
- Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11), 976-989. doi:10.1016/1044-0305(94)80016-2
- Fan, Y., Wang, J., Xu, Y., Wang, Y., Song, T., Liang, X., Jin, F., & Su, D. (2021). Anti-Warburg effect by targeting HRD1-PFKP pathway may inhibit breast cancer progression. *Cell Commun Signal*, 19(1), 18. doi:10.1186/s12964-020-00679-7
- Fang, X., Lloyd, C. J., & Palsson, B. O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 18(12), 731-743. doi:10.1038/s41579-020-00440-4
- Faria, S. S., Morris, C. F. M., Silva, A. R., Fonseca, M. P., Forget, P., Castro, M. S., & Fontes, W. (2017). A Timely Shift from Shotgun to Targeted Proteomics and How It Can Be Groundbreaking for Cancer Research. *Frontiers in Oncology*, 7(13). doi:10.3389/fonc.2017.00013
- Feng, D., Witkowski, A., & Smith, S. (2009). Down-regulation of mitochondrial acyl carrier protein in mammalian cells compromises protein lipoylation and respiratory complex I and results in cell death. *J Biol Chem*, 284(17), 11436-11445. doi:10.1074/jbc.M806991200
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926), 64. doi:10.1126/science.2675315
- Fernando, R., & Smith, T. J. (2020). Slit2 regulates hyaluronan & cytokine synthesis in fibrocytes: Potential relevance to thyroid associated ophthalmopathy. *J Clin Endocrinol Metab*. doi:10.1210/clinem/dgaa684
- Fernando, R., & Smith, T. J. (2021). Slit2 Regulates Hyaluronan & Cytokine Synthesis in Fibrocytes: Potential Relevance to Thyroid-Associated Ophthalmopathy. *J Clin Endocrinol Metab*, 106(1), e20-e33. doi:10.1210/clinem/dgaa684
- Ferrer, C. M., Lynch, T. P., Sodi, V. L., Falcone, J. N., Schwab, L. P., Peacock, D. L., Voadlo, D. J., Seagroves, T. N., & Reginato, M. J. (2014). O-GlcNAcylation regulates cancer metabolism and survival stress signaling via regulation of the HIF-1



pathway. *Mol Cell*, 54(5), 820-831.  
doi:10.1016/j.molcel.2014.04.026

- Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T., Choi, H., El Rayes, T., Ryu, S., Troeger, J., Schwabe, R. F., Vahdat, L. T., Altorki, N. K., Mittal, V., & Gao, D. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*, 527(7579), 472-476. doi:10.1038/nature15748
- Flynn, P., Wongdagger, M., Zavar, M., Dean, N. M., & Stokoe, D. (2000). Inhibition of PDK-1 activity causes a reduction in cell proliferation and survival. *Curr Biol*, 10(22), 1439-1442. doi:10.1016/s0960-9822(00)00801-0
- Fougner, C., Bergholtz, H., Norum, J. H., & Sorlie, T. (2020). Re-definition of claudin-low as a breast cancer phenotype. *Nat Commun*, 11(1), 1787. doi:10.1038/s41467-020-15574-5
- Foulkes, W. D., Smith, I. E., & Reis-Filho, J. S. (2010). Triple-Negative Breast Cancer. *New England Journal of Medicine*, 363(20), 1938-1948. doi:10.1056/NEJMra1001389
- Franz-Wachtel, M., Eisler, S. A., Krug, K., Wahl, S., Carpy, A., Nordheim, A., Pfizenmaier, K., Hausser, A., & Macek, B. (2012). Global detection of protein kinase D-dependent phosphorylation events in nocodazole-treated human cells. *Mol Cell Proteomics*, 11(5), 160-170. doi:10.1074/mcp.M111.016014
- Friedl, P., Locker, J., Sahai, E., & Segall, J. E. (2012). Classifying collective cancer cell invasion. *Nat Cell Biol*, 14(8), 777-783. doi:10.1038/ncb2548
- Friso, S., Udali, S., De Santis, D., & Choi, S. W. (2017). One-carbon metabolism and epigenetics. *Mol Aspects Med*, 54, 28-36. doi:10.1016/j.mam.2016.11.007
- Frixen, U. H., Behrens, J., Sachs, M., Eberle, G., Voss, B., Warda, A., Löchner, D., & Birchmeier, W. (1991). E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J Cell Biol*, 113(1), 173-185. doi:10.1083/jcb.113.1.173
- Fujita, M., Imadome, K., Somasundaram, V., Kawanishi, M., Karasawa, K., & Wink, D. A. (2020). Metabolic characterization of aggressive breast cancer cells exhibiting invasive phenotype: impact of non-cytotoxic doses of 2-DG on diminishing invasiveness. *BMC Cancer*, 20(1), 929. doi:10.1186/s12885-020-07414-y

- Gabel, G., Northoff, B. H., Weinzierl, I., Ludwig, S., Hinterseher, I., Wilfert, W., Teupser, D., Doderer, S. A., Bergert, H., Schonleben, F., Lindeman, J. H. N., & Holdt, L. M. (2017). Molecular Fingerprint for Terminal Abdominal Aortic Aneurysm Disease. *J Am Heart Assoc*, 6(12). doi:10.1161/jaha.117.006798
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), 3718-3720. doi:10.1093/bioinformatics/btv428
- Gallazzini, M., & Burg, M. B. (2009). What's new about osmotic regulation of glycerophosphocholine. *Physiology (Bethesda)*, 24, 245-249. doi:10.1152/physiol.00009.2009
- Gaude, E., & Frezza, C. (2014). Defects in mitochondrial metabolism and cancer. *Cancer Metab*, 2, 10. doi:10.1186/2049-3002-2-10
- Gaude, E., & Frezza, C. (2016). Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nature Communications*, 7(1), 13041. doi:10.1038/ncomms13041
- Ghandi, M., Huang, F. W., Jane-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., 3rd, Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paolella, B. R., Lawrence, M. S., Akbani, R., Lu, Y., Tiv, H. L., Gokhale, P. C., de Weck, A., Mansour, A. A., Oh, C., Shih, J., Hadi, K., Rosen, Y., Bistline, J., Venkatesan, K., Reddy, A., Sonkin, D., Liu, M., Lehar, J., Korn, J. M., Porter, D. A., Jones, M. D., Golji, J., Caponigro, G., Taylor, J. E., Dunning, C. M., Creech, A. L., Warren, A. C., McFarland, J. M., Zamanighomi, M., Kauffmann, A., Stransky, N., Imielinski, M., Maruvka, Y. E., Cherniack, A. D., Tsherniak, A., Vazquez, F., Jaffe, J. D., Lane, A. A., Weinstock, D. M., Johannessen, C. M., Morrissey, M. P., Stegmeier, F., Schlegel, R., Hahn, W. C., Getz, G., Mills, G. B., Boehm, J. S., Golub, T. R., Garraway, L. A., & Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 569(7757), 503-508. doi:10.1038/s41586-019-1186-3
- Giannoni, E., Parri, M., & Chiarugi, P. (2012). EMT and oxidative stress: a bidirectional interplay affecting tumor malignancy. *Antioxid Redox Signal*, 16(11), 1248-1263. doi:10.1089/ars.2011.4280
- Giskeødegård, G. F., Grinde, M. T., Sitter, B., Axelson, D. E., Lundgren, S., Fjøsne, H. E., Dahl, S., Gribbestad, I. S., & Bathen, T. F. (2010). Multivariate modeling and prediction of

- breast cancer prognostic factors using MR metabolomics. *J Proteome Res*, 9(2), 972-979. doi:10.1021/pr9008783
- Glunde, K., Jie, C., & Bhujwalla, Z. M. (2006). Mechanisms of indomethacin-induced alterations in the choline phospholipid metabolism of breast cancer cells. *Neoplasia*, 8(9), 758-771. doi:10.1593/neo.06187
- Glunde, K., Bhujwalla, Z. M., & Ronen, S. M. (2011). Choline metabolism in malignant transformation. *Nat Rev Cancer*, 11(12), 835-848. doi:10.1038/nrc3162
- Golkowski, M., Lau, H. T., Chan, M., Kenerson, H., Vidadala, V. N., Shoemaker, A., Maly, D. J., Yeung, R. S., Gujral, T. S., & Ong, S. E. (2020). Pharmacoproteomics Identifies Kinase Pathways that Drive the Epithelial-Mesenchymal Transition and Drug Resistance in Hepatocellular Carcinoma. *Cell Syst*, 11(2), 196-207.e197. doi:10.1016/j.cels.2020.07.006
- Goodwin, C. R., Ahmed, A. K., & Xia, S. (2019). UDP- $\alpha$ -D-glucose 6-dehydrogenase: a promising target for glioblastoma. *Oncotarget*, 10(16), 1542-1543. doi:10.18632/oncotarget.26670
- Gopal, S. K., Greening, D. W., Mathias, R. A., Ji, H., Rai, A., Chen, M., Zhu, H. J., & Simpson, R. J. (2015). YBX1/YB-1 induces partial EMT and tumourigenicity through secretion of angiogenic factors into the extracellular microenvironment. *Oncotarget*, 6(15), 13718-13730. doi:10.18632/oncotarget.3764
- Greenburg, G., & Hay, E. D. (1982). Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol*, 95(1), 333-339. doi:10.1083/jcb.95.1.333
- Greenburg, G., & Hay, E. D. (1988). Cytoskeleton and thyroglobulin expression change during transformation of thyroid epithelium to mesenchyme-like cells. *Development (Cambridge, England)*, 102(3), 605-622.
- Grinde, M. T., Skrbo, N., Moestue, S. A., Rødland, E. A., Borgan, E., Kristian, A., Sitter, B., Bathen, T. F., Børresen-Dale, A.-L., Mælandsmo, G. M., Engebraaten, O., Sørli, T., Marangoni, E., & Gribbestad, I. S. (2014). Interplay of choline metabolites and genes in patient-derived breast cancer xenografts. *Breast Cancer Research*, 16(1), R5. doi:10.1186/bcr3597
- Gröger, C. J., Grubinger, M., Waldhör, T., Vierlinger, K., & Mikulits, W. (2012). Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer

- progression. *PLoS One*, 7(12), e51136. doi:10.1371/journal.pone.0051136
- Gu, Y., Mi, W., Ge, Y., Liu, H., Fan, Q., Han, C., Yang, J., Han, F., Lu, X., & Yu, W. (2010). GlcNAcylation plays an essential role in breast cancer metastasis. *Cancer Res*, 70(15), 6344-6351. doi:10.1158/0008-5472.CAN-09-1887
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847-2849. doi:10.1093/bioinformatics/btw313
- Gudjonsson, T., Villadsen, R., Nielsen, H. L., Rønnov-Jessen, L., Bissell, M. J., & Petersen, O. W. (2002). Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes & Development*, 16(6), 693-706. doi:10.1101/gad.952602
- Guo, L., Peng, W., Tao, J., Lan, Z., Hei, H., Tian, L., Pan, W., Wang, L., & Zhang, X. (2016). Hydrogen Sulfide Inhibits Transforming Growth Factor-beta1-Induced EMT via Wnt/Catenin Pathway. *PLoS One*, 11(1), e0147018. doi:10.1371/journal.pone.0147018
- Györfy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., & Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*, 123(3), 725-731. doi:10.1007/s10549-009-0674-9
- Ha, N. T., & Lee, C. H. (2020). Roles of Farnesyl-Diphosphate Farnesyltransferase 1 in Tumour and Tumour Microenvironments. *Cells*, 9(11). doi:10.3390/cells9112352
- Hagiuda, D., Nagashio, R., Ichinoe, M., Tsuchiya, B., Igawa, S., Naoki, K., Satoh, Y., Murakumo, Y., Saegusa, M., & Sato, Y. (2019). Clinicopathological and prognostic significance of nuclear UGDH localization in lung adenocarcinoma. *Biomed Res*, 40(1), 17-27. doi:10.2220/biomedres.40.17
- Halestrap, A. P. (2013). The SLC16 gene family - structure, role and regulation in health and disease. *Mol Aspects Med*, 34(2-3), 337-349. doi:10.1016/j.mam.2012.05.003
- Halldorsson, S., Rohatgi, N., Magnusdottir, M., Choudhary, K. S., Gudjonsson, T., Knutsen, E., Barkovskaya, A., Hilmarsson, B., Perander, M., Maelandsmo, G. M., Gudmundsson, S., & Rolfsson, O. (2017). Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal

- transition. *Cancer Lett*, 396, 117-129.  
doi:10.1016/j.canlet.2017.03.019
- Hay, E. (1968). *Organization and fine structure of epithelium and mesenchyme in the developing chick embryo*.
- Hayes, J. D., Dinkova-Kostova, A. T., & Tew, K. D. (2020). Oxidative Stress in Cancer. *Cancer Cell*, 38(2), 167-197.  
doi:10.1016/j.ccell.2020.06.001
- Herschkwitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., Backlund, M. G., Yin, Y., Khramtsov, A. I., Bastein, R., Quackenbush, J., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, L., Furth, P. A., Palazzo, J. P., Olopade, O. I., Bernard, P. S., Churchill, G. A., Van Dyke, T., & Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, 8(5), R76.  
doi:10.1186/gb-2007-8-5-r76
- Hill, J. J., Tremblay, T. L., Cantin, C., O'Connor-McCourt, M., Kelly, J. F., & Lenferink, A. E. (2009). Glycoproteomic analysis of two mouse mammary cell lines during transforming growth factor (TGF)-beta induced epithelial to mesenchymal transition. *Proteome Sci*, 7, 2. doi:10.1186/1477-5956-7-2
- Hilmarsdóttir, B., Briem, E., Sigurdsson, V., Franzdóttir, S. R., Ringnér, M., Arason, A. J., Bergthorsson, J. T., Magnusson, M. K., & Gudjonsson, T. (2015). MicroRNA-200c-141 and  $\Delta$ Np63 are required for breast epithelial differentiation and branching morphogenesis. *Developmental Biology*, 403(2), 150-161.  
doi:10.1016/j.ydbio.2015.05.007
- Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., Malta, T. M., Stuart, J. M., Benz, C. C., & Laird, P. W. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*, 173(2), 291-304.e296.  
doi:10.1016/j.cell.2018.03.022
- Hofmarcher, T., Lindgren, P., Wilking, N., & Jonsson, B. (2020). The cost of cancer in Europe 2018. *Eur J Cancer*, 129, 41-49.  
doi:10.1016/j.ejca.2020.01.011
- Horn, M., Ueckert, T., Fritzeimer, K., Tham, K., Paschke, C., Berg, F., Pfaff, H., Jiang, X., Li, S., & Lopez-Ferrer, D. (2016). New

Method for Label-Free Quantification in the Proteome Discoverer Framework.

- Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*, *43*(Database issue), D512-520. doi:10.1093/nar/gku1267
- Hou, J. M., Krebs, M., Ward, T., Sloane, R., Priest, L., Hughes, A., Clack, G., Ranson, M., Blackhall, F., & Dive, C. (2011). Circulating tumor cells as a window on metastasis biology in lung cancer. *Am J Pathol*, *178*(3), 989-996. doi:10.1016/j.ajpath.2010.12.003
- Hou, T., Zhang, R., Jian, C., Ding, W., Wang, Y., Ling, S., Ma, Q., Hu, X., Cheng, H., & Wang, X. (2019). NDUFB1 confers cardio-protection by enhancing mitochondrial bioenergetics through coordination of respiratory complex and supercomplex assembly. *Cell Res*, *29*(9), 754-766. doi:10.1038/s41422-019-0208-x
- Hoxhaj, G., & Manning, B. D. (2020). The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. *Nature reviews. Cancer*, *20*(2), 74-88. doi:10.1038/s41568-019-0216-7
- Hsu, A. H., Lum, M. A., Shim, K. S., Frederick, P. J., Morrison, C. D., Chen, B., Lele, S. M., Sheinin, Y. M., Daikoku, T., Dey, S. K., Leone, G., Black, A. R., & Black, J. D. (2018). Crosstalk between PKC $\alpha$  and PI3K/AKT Signaling Is Tumor Suppressive in the Endometrium. *Cell Rep*, *24*(3), 655-669. doi:10.1016/j.celrep.2018.06.067
- Hu, Y., Riesland, L., Paterson, A. J., & Kudlow, J. E. (2004). Phosphorylation of mouse glutamine-fructose-6-phosphate amidotransferase 2 (GFAT2) by cAMP-dependent protein kinase increases the enzyme activity. *J Biol Chem*, *279*(29), 29988-29993. doi:10.1074/jbc.M401547200
- Hua, W., Ten Dijke, P., Kostidis, S., Giera, M., & Hornsveld, M. (2020). TGF $\beta$ -induced metabolic reprogramming during epithelial-to-mesenchymal transition in cancer. *Cell Mol Life Sci*, *77*(11), 2103-2123. doi:10.1007/s00018-019-03398-6
- Huang, D., Casale, G. P., Tian, J., Lele, S. M., Pisarev, V. M., Simpson, M. A., & Hemstreet, G. P., 3rd. (2010). Udp-glucose dehydrogenase as a novel field-specific candidate biomarker of prostate cancer. *Int J Cancer*, *126*(2), 315-327. doi:10.1002/ijc.24820

- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57. doi:10.1038/nprot.2008.211
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1), 1-13. doi:10.1093/nar/gkn923
- Huber, M. A., Azoitei, N., Baumann, B., Grunert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H., & Wirth, T. (2004). NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest*, 114(4), 569-581. doi:10.1172/JCI21358
- Huber, M. A., Azoitei, N., Baumann, B., Grünert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H., & Wirth, T. (2004). NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest*, 114(4), 569-581. doi:10.1172/jci21358
- Hugo, H., Ackland, M. L., Blick, T., Lawrence, M. G., Clements, J. A., Williams, E. D., & Thompson, E. W. (2007). Epithelial-mesenchymal and mesenchymal-epithelial transitions in carcinoma progression. *J Cell Physiol*, 213(2), 374-383. doi:10.1002/jcp.21223
- Hugo, H. J., Gunasinghe, N., Hollier, B. G., Tanaka, T., Blick, T., Toh, A., Hill, P., Gilles, C., Waltham, M., & Thompson, E. W. (2017). Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. *Breast Cancer Res*, 19(1), 86. doi:10.1186/s13058-017-0880-z
- Hutchinson, K. E., Yost, S. E., Chang, C. W., Johnson, R. M., Carr, A. R., McAdam, P. R., Halligan, D. L., Chang, C. C., Schmolze, D., Liang, J., & Yuan, Y. (2020). Comprehensive Profiling of Poor-Risk Paired Primary and Recurrent Triple-Negative Breast Cancers Reveals Immune Phenotype Shifts. *Clin Cancer Res*, 26(3), 657-668. doi:10.1158/1078-0432.Ccr-19-1773
- Hwang, E. Y., Huh, J. W., Choi, M. M., Choi, S. Y., Hong, H. N., & Cho, S. W. (2008). Inhibitory effects of gallic acid and quercetin on UDP-glucose dehydrogenase activity. *FEBS Lett*, 582(27), 3793-3797. doi:10.1016/j.febslet.2008.10.010

- Idelchik, M., Begley, U., Begley, T. J., & Melendez, J. A. (2017). Mitochondrial ROS control of cancer. *Semin Cancer Biol*, *47*, 57-66. doi:10.1016/j.semcancer.2017.04.005
- Inghorsson, S., Andersen, K., Hilmarsdottir, B., Maelandsmo, G. M., Magnusson, M. K., & Gudjonsson, T. (2016). HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. *Oncogene*, *35*(32), 4244-4255. doi:10.1038/onc.2015.489
- Iorio, E., Mezzanzanica, D., Alberti, P., Spadaro, F., Ramoni, C., D'Ascenzo, S., Millimaggi, D., Pavan, A., Dolo, V., Canevari, S., & Podo, F. (2005). Alterations of choline phospholipid metabolism in ovarian tumor progression. *Cancer Res*, *65*(20), 9369-9376. doi:10.1158/0008-5472.CAN-05-1146
- Iorio, E., Caramujo, M. J., Cecchetti, S., Spadaro, F., Carpinelli, G., Canese, R., & Podo, F. (2016). Key Players in Choline Metabolic Reprograming in Triple-Negative Breast Cancer. *Frontiers in Oncology*, *6*(205). doi:10.3389/fonc.2016.00205
- Izzo, V., Bravo-San Pedro, J. M., Sica, V., Kroemer, G., & Galluzzi, L. (2016). Mitochondrial Permeability Transition: New Findings and Persisting Uncertainties. *Trends Cell Biol*, *26*(9), 655-667. doi:10.1016/j.tcb.2016.04.006
- Jackson, M. R., Melideo, S. L., & Jorns, M. S. (2012). Human sulfide:quinone oxidoreductase catalyzes the first step in hydrogen sulfide metabolism and produces a sulfane sulfur metabolite. *Biochemistry*, *51*(34), 6804-6815. doi:10.1021/bi300778t
- James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys*, *30*(4), 279-331. doi:10.1017/s0033583597003399
- Jandova, J., & Wondrak, G. T. (2020). Genomic GLO1 deletion modulates TXNIP expression, glucose metabolism, and redox homeostasis while accelerating human A375 malignant melanoma tumor growth. *Redox Biol*, *39*, 101838. doi:10.1016/j.redox.2020.101838
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., & D'Eustachio, P. (2019). The reactome pathway knowledgebase. *Nucleic Acids Res*, *48*(D1), D498-D503. doi:10.1093/nar/gkz1031



- Jeziarska-Drutel, A., Rosenzweig, S. A., & Neumann, C. A. (2013). Role of oxidative stress and the microenvironment in breast cancer development and progression. *Adv Cancer Res*, *119*, 107-125. doi:10.1016/B978-0-12-407190-2.00003-4
- Jia, D., Li, X., Bocci, F., Tripathi, S., Deng, Y., Jolly, M. K., Onuchic, J. N., & Levine, H. (2019). Quantifying Cancer Epithelial-Mesenchymal Plasticity and its Association with Stemness and Immune Response. *J Clin Med*, *8*(5). doi:10.3390/jcm8050725
- Jiang, J., Tang, Y. L., & Liang, X. H. (2011). EMT: a new vision of hypoxia promoting cancer progression. *Cancer Biol Ther*, *11*(8), 714-723. doi:10.4161/cbt.11.8.15274
- Jiang, L., Shestov, A. A., Swain, P., Yang, C., Parker, S. J., Wang, Q. A., Terada, L. S., Adams, N. D., McCabe, M. T., Pietrak, B., Schmidt, S., Metallo, C. M., Dranka, B. P., Schwartz, B., & DeBerardinis, R. J. (2016). Reductive carboxylation supports redox homeostasis during anchorage-independent growth. *Nature*, *532*(7598), 255-258. doi:10.1038/nature17393
- Jolly, M. K., Somarelli, J. A., Sheth, M., Biddle, A., Tripathi, S. C., Armstrong, A. J., Hanash, S. M., Bapat, S. A., Rangarajan, A., & Levine, H. (2019). Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. *Pharmacol Ther*, *194*, 161-184. doi:10.1016/j.pharmthera.2018.09.007
- Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer*, *16*(9), 582-598. doi:10.1038/nrc.2016.73
- Karvelsson, S. T., Sigurdsson, A., Seip, K., Grinde, M. T., Wang, Q., Johannsson, F., Mælandsmo, G. M., Moestue, S. A., Rolfsson, O., & Halldorsson, S. (2021). EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition. *Molecular Cancer Research*. doi:10.1158/1541-7786.MCR-20-0962
- Karvelsson, S. T., Wang, Q., Hilmarsdottir, B., Sigurdsson, A., Moestue, S. A., Mælandsmo, G. M., Halldorsson, S., Gudmundsson, S., & Rolfsson, O. (2021). Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer. *NPJ Syst Biol Appl*, *7*(1), 36. doi:10.1038/s41540-021-00195-5
- Kil, Y. J., Becker, C., Sandoval, W., Goldberg, D., & Bern, M. (2011). Preview: a program for surveying shotgun proteomics tandem mass spectrometry data. *Anal Chem*, *83*(13), 5259-5267. doi:10.1021/ac200609a

- Kim, H. S., Mendiratta, S., Kim, J., Pecot, C. V., Larsen, J. E., Zubovych, I., Seo, B. Y., Kim, J., Eskiocak, B., Chung, H., McMillan, E., Wu, S., De Brabander, J., Komurov, K., Toombs, J. E., Wei, S., Peyton, M., Williams, N., Gazdar, A. F., Posner, B. A., Brekken, R. A., Sood, A. K., Deberardinis, R. J., Roth, M. G., Minna, J. D., & White, M. A. (2013). Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer. *Cell*, *155*(3), 552-566. doi:10.1016/j.cell.2013.09.041
- Kim, J., Lee, H. M., Cai, F., Ko, B., Yang, C., Lieu, E. L., Muhammad, N., Rhyne, S., Li, K., Haloul, M., Gu, W., Faubert, B., Kaushik, A. K., Cai, L., Kasiri, S., Marriam, U., Nham, K., Girard, L., Wang, H., Sun, X., Kim, J., Minna, J. D., Unsal-Kacmaz, K., & DeBerardinis, R. J. (2020). The hexosamine biosynthesis pathway is a targetable liability in KRAS/LKB1 mutant lung cancer. *Nature Metabolism*, *2*(12), 1401-1412. doi:10.1038/s42255-020-00316-0
- Kim, S., Kim, D. H., Jung, W. H., & Koo, J. S. (2013). Expression of glutamine metabolism-related proteins according to molecular subtype of breast cancer. *Endocr Relat Cancer*, *20*(3), 339-348. doi:10.1530/erc-12-0398
- Koch, K., Hartmann, R., Schröter, F., Suwala, A. K., Maciaczyk, D., Krüger, A. C., Willbold, D., Kahlert, U. D., & Maciaczyk, J. (2016). Reciprocal regulation of the cholinic phenotype and epithelial-mesenchymal transition in glioblastoma cells. *Oncotarget*, *7*(45), 73414-73431. doi:10.18632/oncotarget.12337
- Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., Vidović, D., Forlin, M., Kelley, T. T., D'Urso, A., Allen, B. K., Torre, D., Jagodnik, K. M., Wang, L., Jenkins, S. L., Mader, C., Niu, W., Fazel, M., Mahi, N., Pilarczyk, M., Clark, N., Shamsaei, B., Meller, J., Vasiliauskas, J., Reichard, J., Medvedovic, M., Ma'ayan, A., Pillai, A., & Schürer, S. C. (2017). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*, *46*(D1), D558-D566. doi:10.1093/nar/gkx1063
- Köllermann, J., Schlomm, T., Bang, H., Schwall, G. P., von Eichel-Streiber, C., Simon, R., Schostak, M., Huland, H., Berg, W., Sauter, G., Klocker, H., & Schrattenholz, A. (2008). Expression and prognostic relevance of annexin A3 in prostate cancer. *Eur Urol*, *54*(6), 1314-1323. doi:10.1016/j.eururo.2008.01.001

- Krebs, A. M., Mitschke, J., Lasierra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., Boettcher, M., Mouggiakakos, D., Reichardt, W., Bronsert, P., Brunton, V. G., Pilarsky, C., Winkler, T. H., Brabletz, S., Stemmler, M. P., & Brabletz, T. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat Cell Biol*, *19*(5), 518-529. doi:10.1038/ncb3513
- Kresovich, J. K., Zheng, Y., Cardenas, A., Joyce, B. T., Rifas-Shiman, S. L., Oken, E., Gillman, M. W., Hivert, M. F., Baccarelli, A. A., & Hou, L. (2017). Cord blood DNA methylation and adiposity measures in early and mid-childhood. *Clin Epigenetics*, *9*, 86. doi:10.1186/s13148-017-0384-9
- Kröger, C., Afeyan, A., Mraz, J., Eaton, E. N., Reinhardt, F., Khodor, Y. L., Thiru, P., Bierie, B., Ye, X., Burge, C. B., & Weinberg, R. A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci U S A*, *116*(15), 7353-7362. doi:10.1073/pnas.1812876116
- Kuang, S. Q., Tong, W. G., Yang, H., Lin, W., Lee, M. K., Fang, Z. H., Wei, Y., Jelinek, J., Issa, J. P., & Garcia-Manero, G. (2008). Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. *Leukemia*, *22*(8), 1529-1538. doi:10.1038/leu.2008.130
- Labuschagne, C. F., Cheung, E. C., Blagih, J., Domart, M. C., & Vousden, K. H. (2019). Cell Clustering Promotes a Metabolic Switch that Supports Metastatic Colonization. *Cell Metab*. doi:10.1016/j.cmet.2019.07.014
- Lachat, C., Peixoto, P., & Hervouet, E. (2021). Epithelial to Mesenchymal Transition History: From Embryonic Development to Cancers. *Biomolecules*, *11*(6). doi:10.3390/biom11060782
- Lambert, A. W., & Weinberg, R. A. (2021). Linking EMT programmes to normal and neoplastic epithelial stem cells. *Nat Rev Cancer*, *21*(5), 325-338. doi:10.1038/s41568-021-00332-6
- Lau, C. E., Tredwell, G. D., Ellis, J. K., Lam, E. W., & Keun, H. C. (2017). Metabolomic characterisation of the effects of oncogenic PIK3CA transformation in a breast epithelial cell line. *Sci Rep*, *7*, 46079. doi:10.1038/srep46079
- Lawrence, R. T., Perez, E. M., Hernandez, D., Miller, C. P., Haas, K. M., Irie, H. Y., Lee, S. I., Blau, C. A., & Villen, J. (2015). The proteomic landscape of triple-negative breast cancer. *Cell Rep*, *11*(4), 630-644. doi:10.1016/j.celrep.2015.03.050

- Lawton, B. R., Sosa, J. A., Roman, S., & Krause, D. S. (2013). Effect of a matrigel sandwich on endodermal differentiation of human embryonic stem cells. *Stem Cell Rev*, 9(5), 578-585. doi:10.1007/s12015-013-9447-2
- Lebret, S. C., Newgreen, D. F., Thompson, E. W., & Ackland, M. L. (2007). Induction of epithelial to mesenchymal transition in PMC42-LA human breast carcinoma cells by carcinoma-associated fibroblast secreted factors. *Breast Cancer Res*, 9(1), R19. doi:10.1186/bcr1656
- Li, D., Liu, Z., Ding, X., & Qin, Z. (2021). AEBP1 Is One of the Epithelial-Mesenchymal Transition Regulatory Genes in Colon Adenocarcinoma. *Biomed Res Int*, 2021, 3108933. doi:10.1155/2021/3108933
- Li, H., Zhong, A., Li, S., Meng, X., Wang, X., Xu, F., & Lai, M. (2017). The integrated pathway of TGF $\beta$ /Snail with TNF $\alpha$ /NF $\kappa$ B may facilitate the tumor-stroma interaction in the EMT process and colorectal cancer prognosis. *Scientific Reports*, 7(1), 4915. doi:10.1038/s41598-017-05280-6
- Li, H., Ning, S., Ghandi, M., Kryukov, G. V., Gopal, S., Deik, A., Souza, A., Pierce, K., Keskula, P., Hernandez, D., Ann, J., Shkoza, D., Apfel, V., Zou, Y., Vazquez, F., Barretina, J., Pagliarini, R. A., Galli, G. G., Root, D. E., Hahn, W. C., Tsherniak, A., Giannakis, M., Schreiber, S. L., Clish, C. B., Garraway, L. A., & Sellers, W. R. (2019). The landscape of cancer cell line metabolism. *Nat Med*, 25(5), 850-860. doi:10.1038/s41591-019-0404-8
- Li, J., Condello, S., Thomes-Pepin, J., Ma, X., Xia, Y., Hurley, T. D., Matei, D., & Cheng, J. X. (2017). Lipid Desaturation Is a Metabolic Marker and Therapeutic Target of Ovarian Cancer Stem Cells. *Cell Stem Cell*, 20(3), 303-314.e305. doi:10.1016/j.stem.2016.11.004
- Li, J., Ye, Y., Liu, Z., Zhang, G., Dai, H., Li, J., Zhou, B., Li, Y., Zhao, Q., Huang, J., Feng, J., Liu, S., Ruan, P., Wang, J., Liu, J., Huang, M., Liu, X., Yu, S., Liang, Z., Ma, L., Gou, X., Zhang, G., Chen, N., Lu, Y., Di, C., Xia, Q., Pan, J., Feng, R., Cai, Q., & Su, S. (2022). Macrophage mitochondrial fission improves cancer cell phagocytosis induced by therapeutic antibodies and is impaired by glutamine competition. *Nat Cancer*, 3(4), 453-470. doi:10.1038/s43018-022-00354-5
- Li, M., Bu, X., Cai, B., Liang, P., Li, K., Qu, X., & Shen, L. (2019). Biological role of metabolic reprogramming of cancer cells during epithelial-mesenchymal transition (Review). *Oncol Rep*, 41(2), 727-741. doi:10.3892/or.2018.6882

- Li, X., Wang, X., Tan, Z., Chen, S., & Guan, F. (2016). Role of Glycans in Cancer Cells Undergoing Epithelial-Mesenchymal Transition. *Front Oncol*, 6, 33. doi:10.3389/fonc.2016.00033
- Li, X., Wang, M., Gong, T., Lei, X., Hu, T., Tian, M., Ding, F., Ma, F., Chen, H., & Liu, Z. (2020). A S100A14-CCL2/CXCL5 signaling axis drives breast cancer metastasis. *Theranostics*, 10(13), 5687-5703. doi:10.7150/thno.42087
- Liao, T. T., & Yang, M. H. (2020). Hybrid Epithelial/Mesenchymal State in Cancer Metastasis: Clinical Significance and Regulatory Mechanisms. *Cells*, 9(3). doi:10.3390/cells9030623
- Lim, J. K. M., Delaidelli, A., Minaker, S. W., Zhang, H. F., Colovic, M., Yang, H., Negri, G. L., von Karstedt, S., Lockwood, W. W., Schaffer, P., Leprivier, G., & Sorensen, P. H. (2019). Cystine/glutamate antiporter xCT (SLC7A11) facilitates oncogenic RAS transformation by preserving intracellular redox balance. *Proc Natl Acad Sci U S A*, 116(19), 9433-9442. doi:10.1073/pnas.1821323116
- Lim, S. O., Li, C.-W., Xia, W., Lee, H.-H., Chang, S.-S., Shen, J., Hsu, J. L., Raftery, D., Djukovic, D., Gu, H., Chang, W.-C., Wang, H.-L., Chen, M.-L., Huo, L., Chen, C.-H., Wu, Y., Sahin, A., Hanash, S. M., Hortobagyi, G. N., & Hung, M.-C. (2016). EGFR Signaling Enhances Aerobic Glycolysis in Triple-Negative Breast Cancer Cells to Promote Tumor Growth and Immune Escape. *Cancer Research*, 76(5), 1284. doi:10.1158/0008-5472.CAN-15-2478
- Lin, L. H., Chou, H. C., Chang, S. J., Liao, E. C., Tsai, Y. T., Wei, Y. S., Chen, H. Y., Lin, M. W., Wang, Y. S., Chien, Y. A., Yu, X. R., & Chan, H. L. (2020). Targeting UDP-glucose dehydrogenase inhibits ovarian cancer growth and metastasis. *J Cell Mol Med*. doi:10.1111/jcmm.15808
- Liu, J., Gao, L., Zhang, H., Wang, D., Wang, M., Zhu, J., Pang, C., & Wang, C. (2013). Succinate dehydrogenase 5 (SDH5) regulates glycogen synthase kinase 3 $\beta$ - $\beta$ -catenin-mediated lung cancer metastasis. *J Biol Chem*, 288(41), 29965-29973. doi:10.1074/jbc.M113.450106
- Liu, L., Pan, Y., Ren, X., Zeng, Z., Sun, J., Zhou, K., Liang, Y., Wang, F., Yan, Y., Liao, W., Ding, Y., Liu, X., & Liang, L. (2020). GFPT2 promotes metastasis and forms a positive feedback loop with p65 in colorectal cancer. *Am J Cancer Res*, 10(8), 2510-2522.
- Liu, Q.-L., Luo, M., Huang, C., Chen, H.-N., & Zhou, Z.-G. (2021). Epigenetic Regulation of Epithelial to Mesenchymal Transition

in the Cancer Metastatic Cascade: Implications for Cancer Therapy. *Frontiers in Oncology*, 11, 657546-657546. doi:10.3389/fonc.2021.657546

- Loh, C. Y., Chai, J. Y., Tang, T. F., Wong, W. F., Sethi, G., Shanmugam, M. K., Chong, P. P., & Looi, C. Y. (2019). The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges. *Cells*, 8(10). doi:10.3390/cells8101118
- Loibl, S., Poortmans, P., Morrow, M., Denkert, C., & Curigliano, G. (2021). Breast cancer. *The Lancet*, 397(10286), 1750-1769. doi:10.1016/s0140-6736(20)32381-3
- Loo, J. M., Scherl, A., Nguyen, A., Man, F. Y., Weinberg, E., Zeng, Z., Saltz, L., Paty, P. B., & Tavazoie, S. F. (2015). Extracellular metabolic energetics can promote cancer progression. *Cell*, 160(3), 393-406. doi:10.1016/j.cell.2014.12.018
- Lou, T. F., Sethuraman, D., Dospoy, P., Srivastva, P., Kim, H. S., Kim, J., Ma, X., Chen, P. H., Huffman, K. E., Frink, R. E., Larsen, J. E., Lewis, C., Um, S. W., Kim, D. H., Ahn, J. M., DeBerardinis, R. J., White, M. A., Minna, J. D., & Yoo, H. (2016). Cancer-Specific Production of N-Acetylaspartate via NAT8L Overexpression in Non-Small Cell Lung Cancer and Its Potential as a Circulating Biomarker. *Cancer Prev Res (Phila)*, 9(1), 43-52. doi:10.1158/1940-6207.CAPR-14-0287
- Low-Marchelli, J. M., Ardi, V. C., Vizcarra, E. A., van Rooijen, N., Quigley, J. P., & Yang, J. (2013). Twist1 induces CCL2 and recruits macrophages to promote angiogenesis. *Cancer Res*, 73(2), 662-671. doi:10.1158/0008-5472.Can-12-0653
- Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N., & Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A*, 110(45), 18144-18149. doi:10.1073/pnas.1318192110
- Lunetti, P., Di Giacomo, M., Vergara, D., De Domenico, S., Maffia, M., Zara, V., Capobianco, L., & Ferramosca, A. (2019). Metabolic reprogramming in breast cancer results in distinct mitochondrial bioenergetics between luminal and basal subtypes. *FEBS J*, 286(4), 688-709. doi:10.1111/febs.14756
- Lytle, N. K., Barber, A. G., & Reya, T. (2018). Stem cell fate in cancer growth, progression and therapy resistance. *Nat Rev Cancer*, 18(11), 669-680. doi:10.1038/s41568-018-0056-x
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M.,

- Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, *133*(4), 704-715. doi:10.1016/j.cell.2008.03.027
- Mann, M. (2016). Origins of mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol*, *17*(11), 678. doi:10.1038/nrm.2016.135
- Marshall, S., Bacote, V., & Traxinger, R. R. (1991). Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *Journal of Biological Chemistry*, *266*(8), 4706-4712. doi:10.1016/s0021-9258(19)67706-9
- Martinez-Reyes, I., & Chandel, N. S. (2021). Cancer metabolism: looking forward. *Nat Rev Cancer*, *21*(10), 669-680. doi:10.1038/s41568-021-00378-6
- Marx, V. (2019). A dream of single-cell proteomics. *Nat Methods*, *16*(9), 809-812. doi:10.1038/s41592-019-0540-6
- Mathias, R. A., & Simpson, R. J. (2009). Towards understanding epithelial-mesenchymal transition: a proteomics perspective. *Biochim Biophys Acta*, *1794*(9), 1325-1331. doi:10.1016/j.bbapap.2009.05.001
- Mendez, M. G., Kojima, S.-I., & Goldman, R. D. (2010). Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *24*(6), 1838-1851. doi:10.1096/fj.09-151639
- Menendez, J. A., & Lupu, R. (2017). Fatty acid synthase (FASN) as a therapeutic target in breast cancer. *Expert Opin Ther Targets*, *21*(11), 1001-1016. doi:10.1080/14728222.2017.1381087
- Menzel, N., Fischl, W., Hueging, K., Bankwitz, D., Frentzen, A., Haid, S., Gentzsch, J., Kaderali, L., Bartenschlager, R., & Pietschmann, T. (2012). MAP-kinase regulated cytosolic phospholipase A2 activity is essential for production of infectious hepatitis C virus particles. *PLoS Pathog*, *8*(7), e1002829. doi:10.1371/journal.ppat.1002829
- Miah, S., Banks, C. A., Adams, M. K., Florens, L., Lukong, K. E., & Washburn, M. P. (2016). Advancement of mass spectrometry-based proteomics technologies to explore triple negative breast cancer. *Molecular bioSystems*, *13*(1), 42-55. doi:10.1039/c6mb00639f

- Miettinen, P. J., Ebner, R., Lopez, A. R., & Derynck, R. (1994). TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *J Cell Biol*, *127*(6 Pt 2), 2021-2036. doi:10.1083/jcb.127.6.2021
- Millard, P., Letisse, F., Sokol, S., & Portais, J. C. (2012). IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics*, *28*(9), 1294-1296. doi:10.1093/bioinformatics/bts127
- Mirmalek-Sani, S. H., Stokes, P. J., Tare, R. S., Ralph, E. J., Inglis, S., Hanley, N. A., Houghton, F. D., & Oreffo, R. O. (2009). Derivation of a novel undifferentiated human foetal phenotype in serum-free cultures with BMP-2. *J Cell Mol Med*, *13*(9b), 3541-3555. doi:10.1111/j.1582-4934.2009.00742.x
- Miyake, M., Kakimoto, Y., & Sorimachi, M. (1981). A Gas Chromatographic Method for the Determination of N-Acetyl-l-Aspartic Acid, N-Acetyl- $\alpha$ -Aspartylglutamic Acid and  $\beta$ -Citryl-l-Glutamic Acid and Their Distributions in the Brain and Other Organs of Various Species of Animals. *J Neurochem*, *36*(3), 804-819. doi:https://doi.org/10.1111/j.1471-4159.1981.tb01665.x
- Moestue, S. A., Borgan, E., Huuse, E. M., Lindholm, E. M., Sitter, B., Børresen-Dale, A.-L., Engebraaten, O., Mælandsmo, G. M., & Gribbestad, I. S. (2010). Distinct choline metabolic profiles are associated with differences in gene expression for basal-like and luminal-like breast cancer xenograft models. *BMC Cancer*, *10*(1), 433. doi:10.1186/1471-2407-10-433
- Moestue, S. A., Giskeodegard, G. F., Cao, M. D., Bathen, T. F., & Gribbestad, I. S. (2012). Glycerophosphocholine (GPC) is a poorly understood biomarker in breast cancer. *Proc Natl Acad Sci U S A*, *109*(38), E2506; author reply E2507. doi:10.1073/pnas.1208226109
- Moestue, S. A., Dam, C. G., Gorad, S. S., Kristian, A., Bofin, A., Mælandsmo, G. M., Engebråten, O., Gribbestad, I. S., & Bjørkøy, G. (2013). Metabolic biomarkers for response to PI3K inhibition in basal-like breast cancer. *Breast Cancer Research*, *15*(1), R16. doi:10.1186/bcr3391
- Morandi, A., Taddei, M. L., Chiarugi, P., & Giannoni, E. (2017). Targeting the Metabolic Reprogramming That Controls Epithelial-to-Mesenchymal Transition in Aggressive Tumors. *Front Oncol*, *7*, 40. doi:10.3389/fonc.2017.00040
- Morera, E., Steinhäuser, S. S., Budkova, Z., Ingthorsson, S., Krickler, J., Krueger, A., Traustadottir, G. A., & Gudjonsson, T. (2019).



- YKL-40/CHI3L1 facilitates migration and invasion in HER2 overexpressing breast epithelial progenitor cells and generates a niche for capillary-like network formation. *In Vitro Cellular & Developmental Biology - Animal*. doi:10.1007/s11626-019-00403-x
- Morla, S. (2019). Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation. *Int J Mol Sci*, 20(8). doi:10.3390/ijms20081963
- Mundhenke, C., Meyer, K., Drew, S., & Friedl, A. (2002). Heparan sulfate proteoglycans as regulators of fibroblast growth factor-2 receptor binding in breast carcinomas. *Am J Pathol*, 160(1), 185-194. doi:10.1016/s0002-9440(10)64362-3
- Munjal, A., Hannezo, E., Tsai, T. Y., Mitchison, T. J., & Megason, S. G. (2021). Extracellular hyaluronate pressure shaped by cellular tethers drives tissue morphogenesis. *Cell*, 184(26), 6313-6325.e6318. doi:10.1016/j.cell.2021.11.025
- Muzio, G., Maggiora, M., Paiuzzi, E., Oraldi, M., & Canuto, R. A. (2012). Aldehyde dehydrogenases and cell proliferation. *Free Radical Biology and Medicine*, 52(4), 735-746. doi:https://doi.org/10.1016/j.freeradbiomed.2011.11.033
- Nabeebaccus, A. A., Verma, S., Zoccarato, A., Emanuelli, G., Santos, C. X., Streckfuss-Bomeke, K., & Shah, A. M. (2021). Cardiomyocyte protein O-GlcNAcylation is regulated by GFAT1 not GFAT2. *Biochem Biophys Res Commun*, 583, 121-127. doi:10.1016/j.bbrc.2021.10.056
- Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., & Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*, 7, 548. doi:10.1038/msb.2011.81
- Naidu, S., Shi, L., Magee, P., Middleton, J. D., Laganá, A., Sahoo, S., Leong, H. S., Galvin, M., Frese, K., Dive, C., Guzzardo, V., Fassan, M., & Garofalo, M. (2017). PDGFR-modulated miR-23b cluster and miR-125a-5p suppress lung tumorigenesis by targeting multiple components of KRAS and NF-kB pathways. *Sci Rep*, 7(1), 15441. doi:10.1038/s41598-017-14843-6
- Nath, A., Li, I., Roberts, L. R., & Chan, C. (2015). Elevated free fatty acid uptake via CD36 promotes epithelial-mesenchymal transition in hepatocellular carcinoma. *Sci Rep*, 5, 14752. doi:10.1038/srep14752
- Navas, T., Kinders, R. J., Lawrence, S. M., Ferry-Galow, K. V., Borgel, S., Hollingshead, M. G., Srivastava, A. K., Alcoser, S. Y., Makhlof, H. R., Chuaqui, R., Wilsker, D. F., Konaté, M.

- M., Miller, S. B., Voth, A. R., Chen, L., Vilimas, T., Subramanian, J., Rubinstein, L., Kummar, S., Chen, A. P., Bottaro, D. P., Doroshow, J. H., & Parchment, R. E. (2020). Clinical Evolution of Epithelial-Mesenchymal Transition in Human Carcinomas. *Cancer Res*, *80*(2), 304-318. doi:10.1158/0008-5472.Can-18-3539
- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A., & Gray, J. W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, *10*(6), 515-527. doi:10.1016/j.ccr.2006.10.008
- Newman, A. C., & Maddocks, O. D. K. (2017). One-carbon metabolism in cancer. *Br J Cancer*, *116*(12), 1499-1504. doi:10.1038/bjc.2017.118
- Nieto, M. A., Huang, Ruby Y.-J., Jackson, Rebecca A., & Thiery, Jean P. (2016). EMT: 2016. *Cell*, *166*(1), 21-45. doi:https://doi.org/10.1016/j.cell.2016.06.028
- Niringiyumukiza, J. D., Cai, H., Chen, L., Li, Y., Wang, L., Zhang, M., Xu, X., & Xiang, W. (2019). Protective properties of glycogen synthase kinase-3 inhibition against doxorubicin-induced oxidative damage to mouse ovarian reserve. *Biomed Pharmacother*, *116*, 108963. doi:10.1016/j.biopha.2019.108963
- Nivet, A. L., Vigneault, C., Blondin, P., & Sirard, M. A. (2013). Changes in granulosa cells' gene expression associated with increased oocyte competence in bovine. *Reproduction*, *145*(6), 555-565. doi:10.1530/rep-13-0032
- Ogrodzinski, M. P., Teoh, S. T., & Lunt, S. Y. (2021). Targeting Subtype-Specific Metabolic Preferences in Nucleotide Biosynthesis Inhibits Tumor Growth in a Breast Cancer Model. *Cancer Res*, *81*(2), 303-314. doi:10.1158/0008-5472.CAN-20-1666
- Ohata, T., Yokoo, H., Kamiyama, T., Fukai, M., Aiyama, T., Hatanaka, Y., Hatanaka, K., Wakayama, K., Orimo, T., Kakisaka, T., Kobayashi, N., Matsuno, Y., & Taketomi, A. (2017). Fatty acid-binding protein 5 function in hepatocellular carcinoma through induction of epithelial-mesenchymal transition. *Cancer Med*, *6*(5), 1049-1061. doi:10.1002/cam4.1020

- Oikari, S., Makkonen, K., Deen, A. J., Tyni, I., Karna, R., Tammi, R. H., & Tammi, M. I. (2016). Hexosamine biosynthesis in keratinocytes: roles of GFAT and GNPDA enzymes in the maintenance of UDP-GlcNAc content and hyaluronan synthesis. *Glycobiology*, 26(7), 710-722. doi:10.1093/glycob/cww019
- Oikari, S., Kettunen, T., Tiainen, S., Hayrinen, J., Masarwah, A., Sudah, M., Sutela, A., Vanninen, R., Tammi, M., & Auvinen, P. (2018). UDP-sugar accumulation drives hyaluronan synthesis in breast cancer. *Matrix Biol*, 67, 63-74. doi:10.1016/j.matbio.2017.12.015
- Okazaki, Y., Nakamura, K., Takeda, S., Yoshizawa, I., Yoshida, F., Ohshima, N., Izumi, T., Klein, J. D., Kumrungsee, T., Sands, J. M., & Yanaka, N. (2019). GDE5 inhibition accumulates intracellular glycerophosphocholine and suppresses adipogenesis at a mitotic clonal expansion stage. *Am J Physiol Cell Physiol*, 316(2), C162-C174. doi:10.1152/ajpcell.00305.2018
- Oki, T., Yamazaki, K., Kuromitsu, J., Okada, M., & Tanaka, I. (1999). cDNA Cloning and Mapping of a Novel Subtype of Glutamine:fructose-6-phosphate Amidotransferase (GFAT2) in Human and Mouse. *Genomics*, 57(2), 227-234. doi:https://doi.org/10.1006/geno.1999.5785
- Oliveira, I. A., Allonso, D., Fernandes, T. V. A., Lucena, D. M. S., Ventura, G. T., Dias, W. B., Mohana-Borges, R. S., Pascutti, P. G., & Todeschini, A. R. (2021). Enzymatic and structural properties of human glutamine:fructose-6-phosphate amidotransferase 2 (hGFAT2). *J Biol Chem*, 296, 100180. doi:10.1074/jbc.RA120.015189
- Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., & Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics*, 4(12), 2010-2021. doi:10.1074/mcp.T500030-MCP200
- Ong, S.-E., & Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, 1(6), 2650-2660. doi:10.1038/nprot.2006.427
- Overall, C. M. (2020). The Human Proteome: 90% in the Light, 10% on the Dark Side. *J Proteome Res*, 19(12), 4731-4734. doi:10.1021/acs.jproteome.0c00914

- Oyinlade, O., Wei, S., Lal, B., Lattera, J., Zhu, H., Goodwin, C. R., Wang, S., Ma, D., Wan, J., & Xia, S. (2018). Targeting UDP-alpha-D-glucose 6-dehydrogenase inhibits glioblastoma growth and migration. *Oncogene*, *37*(20), 2615-2629. doi:10.1038/s41388-018-0138-y
- Palma Cde, S., Grassi, M. L., Thomé, C. H., Ferreira, G. A., Albuquerque, D., Pinto, M. T., Ferreira Melo, F. U., Kashima, S., Covas, D. T., Pitteri, S. J., & Faça, V. M. (2016). Proteomic Analysis of Epithelial to Mesenchymal Transition (EMT) Reveals Cross-talk between SNAIL and HDAC1 Proteins in Breast Cancer Cells. *Mol Cell Proteomics*, *15*(3), 906-917. doi:10.1074/mcp.M115.052910
- Pan, Y. R., Vatsyayan, J., Chang, Y. S., & Chang, H. Y. (2008). Epstein-Barr virus latent membrane protein 2A upregulates UDP-glucose dehydrogenase gene expression via ERK and PI3K/Akt pathway. *Cell Microbiol*, *10*(12), 2447-2460. doi:10.1111/j.1462-5822.2008.01221.x
- Panarsky, R., Crooks, D. R., Lane, A. N., Yang, Y., Cassel, T. A., Fan, T. W., Linehan, W. M., & Moscow, J. A. (2020). Fumarate hydratase-deficient renal cell carcinoma cells respond to asparagine by activation of the unfolded protein response and stimulation of the hexosamine biosynthetic pathway. *Cancer Metab*, *8*, 7. doi:10.1186/s40170-020-00214-9
- Pandeswari, P. B., & Sabareesh, V. (2019). Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Advances*, *9*(1), 313-344. doi:10.1039/c8ra07200k
- Park, J. H., Vithayathil, S., Kumar, S., Sung, P. L., Dobrolecki, L. E., Putluri, V., Bhat, V. B., Bhowmik, S. K., Gupta, V., Arora, K., Wu, D., Tsouko, E., Zhang, Y., Maity, S., Donti, T. R., Graham, B. H., Frigo, D. E., Coarfa, C., Yotnda, P., Putluri, N., Sreekumar, A., Lewis, M. T., Creighton, C. J., Wong, L. C., & Kaiparettu, B. A. (2016). Fatty Acid Oxidation-Driven Src Links Mitochondrial Energy Reprogramming and Oncogenic Properties in Triple-Negative Breast Cancer. *Cell Rep*, *14*(9), 2154-2165. doi:10.1016/j.celrep.2016.02.004
- Park, S. Y., Kim, H. S., Kim, N. H., Ji, S., Cha, S. Y., Kang, J. G., Ota, I., Shimada, K., Konishi, N., Nam, H. W., Hong, S. W., Yang, W. H., Roth, J., Yook, J. I., & Cho, J. W. (2010). Snail1 is stabilized by O-GlcNAc modification in hyperglycaemic condition. *The EMBO journal*, *29*(22), 3787-3796. doi:10.1038/emboj.2010.254

- Passi, A., Vigetti, D., Buraschi, S., & Iozzo, R. V. (2019). Dissecting the role of hyaluronan synthases in the tumor microenvironment. *FEBS J*, 286(15), 2937-2949. doi:10.1111/febs.14847
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., De Clercq, S., Minguignon, E., Balsat, C., Sokolow, Y., Dubois, C., De Cock, F., Scozzaro, S., Sopena, F., Lanas, A., D'Haene, N., Salmon, I., Marine, J. C., Voet, T., Sotiropoulou, P. A., & Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nature*, 556(7702), 463-468. doi:10.1038/s41586-018-0040-3
- Pastushenko, I., & Blanpain, C. (2019). EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol*, 29(3), 212-226. doi:10.1016/j.tcb.2018.12.001
- Patra, K. C., & Hay, N. (2014). The pentose phosphate pathway and cancer. *Trends Biochem Sci*, 39(8), 347-354. doi:10.1016/j.tibs.2014.06.005
- Patsialou, A., Wang, Y., Pignatelli, J., Chen, X., Entenberg, D., Oktay, M., & Condeelis, J. S. (2015). Autocrine CSF1R signaling mediates switching between invasion and proliferation downstream of TGFbeta in claudin-low breast tumor cells. *Oncogene*, 34(21), 2721-2731. doi:10.1038/onc.2014.226
- Paul, D., Chanukuppa, V., Reddy, P. J., Taunk, K., Adhav, R., Srivastava, S., Santra, M. K., & Rapole, S. (2016). Global proteomic profiling identifies etoposide chemoresistance markers in non-small cell lung carcinoma. *J Proteomics*, 138, 95-105. doi:10.1016/j.jprot.2016.02.008
- Pavlidis, S., Whitaker-Menezes, D., Castello-Cros, R., Flomenberg, N., Witkiewicz, A. K., Frank, P. G., Casimiro, M. C., Wang, C., Fortina, P., Addya, S., Pestell, R. G., Martinez-Outschoorn, U. E., Sotgia, F., & Lisanti, M. P. (2009). The reverse Warburg effect: aerobic glycolysis in cancer associated fibroblasts and the tumor stroma. *Cell Cycle*, 8(23), 3984-4001. doi:10.4161/cc.8.23.10238
- Pavlova, N. N., & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cell Metab*, 23(1), 27-47. doi:10.1016/j.cmet.2015.12.006
- Pavón, M., Parreño, M., Téllez-Gabriel, M., Sancho, F., López, M., Céspedes, M., Casanova, I., Lopez-Pousa, A., Mangués, M., Quer, M., Barnadas, A., León, X., & Mangués, R. (2012). Gene expression signatures and molecular markers associated with

- clinical outcome in locally advanced head and neck carcinoma. *Carcinogenesis*, *33*(9), 1707-1716. doi:10.1093/carcin/bgs207
- Pei, D., Shu, X., Gassama-Diagne, A., & Thiery, J. P. (2019). Mesenchymal-epithelial transition in development and reprogramming. *Nat Cell Biol*, *21*(1), 44-53. doi:10.1038/s41556-018-0195-z
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., & Vizcaíno, J. A. (2018). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*, *47*(D1), D442-D450. doi:10.1093/nar/gky1106
- Perou, C. M., Sørli, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslén, L. A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A.-L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), 747-752. doi:10.1038/35021093
- Phannasil, P., Thuwajit, C., Warnnissorn, M., Wallace, J. C., MacDonald, M. J., & Jitrapakdee, S. (2015). Pyruvate Carboxylase Is Up-Regulated in Breast Cancer and Essential to Support Growth and Invasion of MDA-MB-231 Cells. *PLoS One*, *10*(6), e0129848. doi:10.1371/journal.pone.0129848
- Phillips, C. M., Zatarain, J. R., Nicholls, M. E., Porter, C., Widen, S. G., Thanki, K., Johnson, P., Jawad, M. U., Moyer, M. P., Randall, J. W., Hellmich, J. L., Maskey, M., Qiu, S., Wood, T. G., Druzhyna, N., Szczesny, B., Módis, K., Szabo, C., Chao, C., & Hellmich, M. R. (2017). Upregulation of Cystathionine-β-Synthase in Colonic Epithelia Reprograms Metabolism and Promotes Carcinogenesis. *Cancer Res*, *77*(21), 5741-5754. doi:10.1158/0008-5472.Can-16-3480
- Pillai, R., Hayashi, M., Zavitsanou, A. M., & Papagiannakopoulos, T. (2022). NRF2: KEAPing Tumors Protected. *Cancer Discov*. doi:10.1158/2159-8290.Cd-21-0922
- Pommier, R. M., Sanlaville, A., Tonon, L., Kielbassa, J., Thomas, E., Ferrari, A., Sertier, A. S., Hollande, F., Martinez, P., Tissier, A., Morel, A. P., Ouzounova, M., & Puisieux, A. (2020). Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat Commun*, *11*(1), 3431. doi:10.1038/s41467-020-17249-7

- Porporato, P. E., Filigheddu, N., Pedro, J. M. B., Kroemer, G., & Galluzzi, L. (2018). Mitochondrial metabolism and cancer. *Cell Res*, 28(3), 265-280. doi:10.1038/cr.2017.155
- Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H. K., Jang, H. G., Jha, A. K., Chen, W. W., Barrett, F. G., Stransky, N., Tsun, Z. Y., Cowley, G. S., Barretina, J., Kalaany, N. Y., Hsu, P. P., Ottina, K., Chan, A. M., Yuan, B., Garraway, L. A., Root, D. E., Mino-Kenudson, M., Brachtel, E. F., Driggers, E. M., & Sabatini, D. M. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature*, 476(7360), 346-350. doi:10.1038/nature10350
- Prabhu, A. V., Luu, W., Sharpe, L. J., & Brown, A. J. (2017). Phosphorylation regulates activity of 7-dehydrocholesterol reductase (DHCR7), a terminal enzyme of cholesterol synthesis. *J Steroid Biochem Mol Biol*, 165(Pt B), 363-368. doi:10.1016/j.jsbmb.2016.08.003
- Prasad, P., Tiwari, A. K., Kumar, K. M., Ammini, A. C., Gupta, A., Gupta, R., & Thelma, B. K. (2010). Association analysis of ADPRT1, AKR1B1, RAGE, GFPT2 and PAI-1 gene polymorphisms with chronic renal insufficiency among Asian Indians with type-2 diabetes. *BMC Med Genet*, 11, 52. doi:10.1186/1471-2350-11-52
- Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X., & Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research*, 12(5), R68. doi:10.1186/bcr2635
- Puisieux, A., Pommier, R. M., Morel, A. P., & Laval, F. (2018). Cellular Pliancy and the Multistep Process of Tumorigenesis. *Cancer Cell*, 33(2), 164-172. doi:10.1016/j.ccell.2018.01.007
- Qiao, D., Skibba, M., Xu, X., Garofalo, R. P., Zhao, Y., & Brasier, A. R. (2021). Paramyxovirus replication induces the Hexosamine Biosynthetic Pathway and Mesenchymal Transition via the IRE1a-XBP1s arm of the Unfolded Protein Response. *Am J Physiol Lung Cell Mol Physiol*. doi:10.1152/ajplung.00127.2021
- Qu, Y., Han, B., Yu, Y., Yao, W., Bose, S., Karlan, B. Y., Giuliano, A. E., & Cui, X. (2015). Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. *PLoS One*, 10(7), e0131285. doi:10.1371/journal.pone.0131285

- Quinzii, C. M., & Lopez, L. C. (2021). Abnormalities of hydrogen sulfide and glutathione pathways in mitochondrial dysfunction. *J Adv Res*, 27, 79-84. doi:10.1016/j.jare.2020.04.002
- Radler, P. D., Wehde, B. L., Triplett, A. A., Shrestha, H., Shepherd, J. H., Pfefferle, A. D., Rui, H., Cardiff, R. D., Perou, C. M., & Wagner, K. U. (2021). Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat Commun*, 12(1), 3742. doi:10.1038/s41467-021-23957-5
- Ramesh, V., Brabletz, T., & Ceppi, P. (2020). Targeting EMT in Cancer with Repurposed Metabolic Inhibitors. *Trends Cancer*, 6(11), 942-950. doi:10.1016/j.trecan.2020.06.005
- Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry*, 75(3), 663-670. doi:10.1021/ac026117i
- Ricciardelli, C., Brooks, J. H., Suwiwat, S., Sakko, A. J., Mayne, K., Raymond, W. A., Seshadri, R., LeBaron, R. G., & Horsfall, D. J. (2002). Regulation of stromal versican expression by breast cancer cells and importance to relapse-free survival in patients with node-negative primary breast cancer. *Clin Cancer Res*, 8(4), 1054-1060.
- Richani, D., Sutton-McDowall, M. L., Frank, L. A., Gilchrist, R. B., & Thompson, J. G. (2014). Effect of epidermal growth factor-like peptides on the metabolism of in vitro- matured mouse oocytes and cumulus cells. *Biol Reprod*, 90(3), 49. doi:10.1095/biolreprod.113.115311
- Riis, S., Murray, J. B., & O'Connor, R. (2020). IGF-1 Signalling Regulates Mitochondria Dynamics and Turnover through a Conserved GSK-3beta-Nrf2-BNIP3 Pathway. *Cells*, 9(1). doi:10.3390/cells9010147
- Rinschen, M. M., Ivanisevic, J., Giera, M., & Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol*, 20(6), 353-367. doi:10.1038/s41580-019-0108-4
- Rodrigues, M. F., Obre, E., de Melo, F. H., Santos, G. C., Jr., Galina, A., Jasiulionis, M. G., Rossignol, R., Rumjanek, F. D., & Amoêdo, N. D. (2016). Enhanced OXPHOS, glutaminolysis and  $\beta$ -oxidation constitute the metastatic phenotype of melanoma cells. *Biochem J*, 473(6), 703-715. doi:10.1042/bj20150645



- Rolfsson, O., Johannsson, F., Magnúsdóttir, M., Paglia, G., Sigurjónsson, O. E., Bordbar, A., Pálsson, S., Brynjólfsson, S., Guethmundsson, S., & Pálsson, B. (2017). Mannose and fructose metabolism in red blood cells during cold storage in SAGM. *Transfusion*, *57*(11), 2665-2676. doi:10.1111/trf.14266
- Roth, L., Srivastava, S., Lindzen, M., Sas-Chen, A., Sheffer, M., Lauriola, M., Euka, Y., Noronha, A., Mancini, M., Lavi, S., Tarcic, G., Pines, G., Nevo, N., Heyman, O., Ziv, T., Rueda, O. M., Gnocchi, D., Pikarsky, E., Admon, A., Caldas, C., & Yarden, Y. (2018). SILAC identifies LAD1 as a filamin-binding regulator of actin dynamics in response to EGF and a marker of aggressive breast tumors. *Sci Signal*, *11*(515), ean0949. doi:10.1126/scisignal.aan0949
- Ruegenberg, S., Horn, M., Pichlo, C., Allmeroth, K., Baumann, U., & Denzel, M. S. (2020). Loss of GFAT-1 feedback regulation activates the hexosamine pathway that modulates protein homeostasis. *Nat Commun*, *11*(1), 687. doi:10.1038/s41467-020-14524-5
- Saberi, B., Shinohara, M., Ybanez, M. D., Hanawa, N., Gaarde, W. A., Kaplowitz, N., & Han, D. (2008). Regulation of H<sub>2</sub>O<sub>2</sub>-induced necrosis by PKC and AMP-activated kinase signaling in primary cultured hepatocytes. *Am J Physiol Cell Physiol*, *295*(1), C50-C63. doi:10.1152/ajpcell.90654.2007
- Schafer, M., Goodenough, S., Moosmann, B., & Behl, C. (2004). Inhibition of glycogen synthase kinase 3 beta is involved in the resistance to oxidative stress in neuronal HT22 cells. *Brain Res*, *1005*(1-2), 84-89. doi:10.1016/j.brainres.2004.01.037
- Schiff, R., Massarweh, S. A., Shou, J., Bharwani, L., Mohsin, S. K., & Osborne, C. K. (2004). Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance. *Clin Cancer Res*, *10*(1 Pt 2), 331s-336s. doi:10.1158/1078-0432.ccr-031212
- Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol*, *21*(12), 729-749. doi:10.1038/s41580-020-00294-x
- Schwab, A., Siddiqui, A., Vazakidou, M. E., Napoli, F., Böttcher, M., Menchicchi, B., Raza, U., Saatci, Ö., Krebs, A. M., Ferrazzi, F., Rapa, I., Dettmer-Wilde, K., Waldner, M. J., Ekici, A. B., Rasheed, S. A. K., Mougiakakos, D., Oefner, P. J., Sahin, O., Volante, M., Greten, F. R., Brabletz, T., & Ceppi, P. (2018). Polyol Pathway Links Glucose Metabolism to the

Aggressiveness of Cancer Cells. *Cancer Res*, 78(7), 1604-1618. doi:10.1158/0008-5472.Can-17-2834

- Schworer, S., Vardhana, S. A., & Thompson, C. B. (2019). Cancer Metabolism Drives a Stromal Regenerative Response. *Cell Metab*, 29(3), 576-591. doi:10.1016/j.cmet.2019.01.015
- Sciacovelli, M., Goncalves, E., Johnson, T. I., Zecchini, V. R., da Costa, A. S., Gaude, E., Drubbel, A. V., Theobald, S. J., Abbo, S. R., Tran, M. G., Rajeeve, V., Cardaci, S., Foster, S., Yun, H., Cutillas, P., Warren, A., Gnanapragasam, V., Gottlieb, E., Franze, K., Huntly, B., Maher, E. R., Maxwell, P. H., Saez-Rodriguez, J., & Frezza, C. (2016). Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature*, 537(7621), 544-547. doi:10.1038/nature19353
- Sciacovelli, M., & Frezza, C. (2017). Metabolic reprogramming and epithelial-to-mesenchymal transition in cancer. *FEBS J*, 284(19), 3132-3144. doi:10.1111/febs.14090
- Scoglio, S., Lo Curcio, V., Catalani, S., Palma, F., Battistelli, S., & Benedetti, S. (2016). Inhibitory effects of Aphanizomenon flos-aquae constituents on human UDP-glucose dehydrogenase activity. *J Enzyme Inhib Med Chem*, 31(6), 1492-1497. doi:10.3109/14756366.2016.1149478
- Sen, S., Kawahara, B., Mahata, S. K., Tsai, R., Yoon, A., Hwang, L., Hu-Moore, K., Villanueva, C., Vajihuddin, A., Parameshwar, P., You, M., Bhaskar, D. L., Gomez, O., Faull, K. F., Farias-Eisner, R., & Chaudhuri, G. (2016). Cystathionine: A novel oncometabolite in human breast cancer. *Arch Biochem Biophys*, 604, 95-102. doi:10.1016/j.abb.2016.06.010
- Sessions, D. T., & Kashatus, D. F. (2021). Mitochondrial dynamics in cancer stem cells. *Cell Mol Life Sci*, 10.1007/s00018-021-03773-2. doi:10.1007/s00018-021-03773-2
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11), 2498-2504. doi:10.1101/gr.1239303
- Shaul, Y. D., Freinkman, E., Comb, W. C., Cantor, J. R., Tam, W. L., Thiru, P., Kim, D., Kanarek, N., Pacold, M. E., Chen, W. W., Bierie, B., Possemato, R., Reinhardt, F., Weinberg, R. A., Yaffe, M. B., & Sabatini, D. M. (2014). Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell*, 158(5), 1094-1109. doi:10.1016/j.cell.2014.07.032

- Shimamura, T., Hsu, T. C., Colburn, N. H., & Bejcek, B. E. (2002). Activation of NF-kappaB is required for PDGF-B chain to transform NIH3T3 cells. *Exp Cell Res*, 274(1), 157-167. doi:10.1006/excr.2001.5449
- Sigurdsson, V., Hilmarsdottir, B., Sigmundsdottir, H., Fridriksdottir, A. J. R., Ringnér, M., Villadsen, R., Borg, A., Agnarsson, B. A., Petersen, O. W., Magnusson, M. K., & Gudjonsson, T. (2011). Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One*, 6(9), e23833. doi:10.1371/journal.pone.0023833
- Simeonov, K. P., Byrns, C. N., Clark, M. L., Norgard, R. J., Martin, B., Stanger, B. Z., Shendure, J., McKenna, A., & Lengner, C. J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell*, 39(8), 1150-1162 e1159. doi:10.1016/j.ccell.2021.05.005
- Simpson, N. E., Tryndyak, V. P., Beland, F. A., & Pogribny, I. P. (2012). An in vitro investigation of metabolically sensitive biomarkers in breast cancer progression. *Breast Cancer Res Treat*, 133(3), 959-968. doi:10.1007/s10549-011-1871-x
- Singh, R. K., Kumar, S., Gautam, P. K., Tomar, M. S., Verma, P. K., Singh, S. P., Kumar, S., & Acharya, A. (2017). Protein kinase C-alpha and the regulation of diverse cell responses. *Biomol Concepts*, 8(3-4), 143-153. doi:10.1515/bmc-2017-0005
- Sinha, D., Saha, P., Samanta, A., & Bishayee, A. (2020). Emerging Concepts of Hybrid Epithelial-to-Mesenchymal Transition in Cancer Progression. *Biomolecules*, 10(11). doi:10.3390/biom10111561
- Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., & Siuzdak, G. (2005). METLIN: a metabolite mass spectral database. *Ther Drug Monit*, 27(6), 747-751. doi:10.1097/01.ftd.0000179845.53213.39
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, 78(3), 779-787. doi:10.1021/ac051437y
- Sonkar, K., Ayyappan, V., Tressler, C. M., Adelaja, O., Cai, R., Cheng, M., & Glunde, K. (2019). Focus on the glycerophosphocholine pathway in choline phospholipid metabolism of cancer. *NMR Biomed*, 32(10), e4112. doi:10.1002/nbm.4112

- Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Jr., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., & Brooks, S. C. (1990). Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res*, *50*(18), 6075-6086.
- Steinhaeuser, S. S., Morera, E., Budkova, Z., Schepsky, A., Wang, Q., Rolfsson, O., Riedel, A., Krueger, A., Hilmarsdottir, B., Maellandsmo, G. M., Valdimarsdottir, B., Sigurdardottir, A. K., Agnarsson, B. A., Jonasson, J. G., Ingthorsson, S., Traustadottir, G. A., Oskarsson, T., & Gudjonsson, T. (2020). ECM1 secreted by HER2-overexpressing breast cancer cells promotes formation of a vascular niche accelerating cancer cell migration and invasion. *Laboratory Investigation*, 1-17. doi:10.1038/s41374-020-0415-6
- Stemmler, M. P., Eccles, R. L., Brabletz, S., & Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. *Nat Cell Biol*, *21*(1), 102-112. doi:10.1038/s41556-018-0196-y
- Stewart, J. D., Marchan, R., Lesjak, M. S., Lambert, J., Hergenroeder, R., Ellis, J. K., Lau, C. H., Keun, H. C., Schmitz, G., Schiller, J., Eibisch, M., Hedberg, C., Waldmann, H., Lausch, E., Tanner, B., Sehoul, J., Sagemueller, J., Staude, H., Steiner, E., & Hengstler, J. G. (2012). Choline-releasing glycerophosphodiesterase EDI3 drives tumor cell migration and metastasis. *Proc Natl Acad Sci U S A*, *109*(21), 8155-8160. doi:10.1073/pnas.1117654109
- Strominger, J. L., Maxwell, E. S., Axelrod, J., & Kalckar, H. M. (1957). Enzymatic Formation of Uridine Diphosphoglucuronic Acid. *Journal of Biological Chemistry*, *224*(1), 79-90. doi:10.1016/s0021-9258(18)65012-4
- Su, J. S., Woods, S. M., & Ronen, S. M. (2012). Metabolic consequences of treatment with AKT inhibitor perifosine in breast cancer cells. *NMR Biomed*, *25*(2), 379-388. doi:10.1002/nbm.1764
- Sun, Y., Liu, X., Zhang, Q., Mao, X., Feng, L., Su, P., Chen, H., Guo, Y., & Jin, F. (2016). Oncogenic potential of TSTA3 in breast cancer and its regulation by the tumor suppressors miR-125a-5p and miR-125b. *Tumour Biol*, *37*(4), 4963-4972. doi:10.1007/s13277-015-4178-4
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, *71*(3), 209-249. doi:10.3322/caac.21660

- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-D613. doi:10.1093/nar/gky1131
- Szymura, S. J., Zaemes, J. P., Allison, D. F., Clift, S. H., D'Innocenzi, J. M., Gray, L. G., McKenna, B. D., Morris, B. B., Bekiranov, S., LeGallo, R. D., Jones, D. R., & Mayo, M. W. (2019). NF-kappaB upregulates glutamine-fructose-6-phosphate transaminase 2 to promote migration in non-small cell lung cancer. *Cell Commun Signal*, *17*(1), 24. doi:10.1186/s12964-019-0335-5
- Tam, W. L., Lu, H., Buikhuisen, J., Soh, B. S., Lim, E., Reinhardt, F., Wu, Z. J., Krall, J. A., Bierie, B., Guo, W., Chen, X., Liu, X. S., Brown, M., Lim, B., & Weinberg, R. A. (2013). Protein kinase C alpha is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell*, *24*(3), 347-364. doi:10.1016/j.ccr.2013.08.005
- Taparra, K., Tran, P. T., & Zachara, N. E. (2016). Hijacking the Hexosamine Biosynthetic Pathway to Promote EMT-Mediated Neoplastic Phenotypes. *Front Oncol*, *6*, 85. doi:10.3389/fonc.2016.00085
- Taparra, K., Wang, H., Malek, R., Lafargue, A., Barbhuiya, M. A., Wang, X., Simons, B. W., Ballew, M., Nugent, K., Groves, J., Williams, R. D., Shiraishi, T., Verdone, J., Yildirim, G., Henry, R., Zhang, B., Wong, J., Wang, K. K.-H., Nelkin, B. D., Pienta, K. J., Felsher, D., Zachara, N. E., & Tran, P. T. (2019). O-GlcNAcylation is required for mutant KRAS-induced lung tumorigenesis. *J Clin Invest*, *128*(11), 4924-4937. doi:10.1172/JCI94844
- Tasdogan, A., Faubert, B., Ramesh, V., Ubellacker, J. M., Shen, B., Solmonson, A., Murphy, M. M., Gu, Z., Gu, W., Martin, M., Kasitinon, S. Y., Vandergriff, T., Mathews, T. P., Zhao, Z., Schadendorf, D., DeBerardinis, R. J., & Morrison, S. J. (2020). Metabolic heterogeneity confers differences in melanoma metastatic potential. *Nature*, *577*(7788), 115-120. doi:10.1038/s41586-019-1847-2
- Tautenhahn, R., Böttcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, *9*(1), 504. doi:10.1186/1471-2105-9-504

- Teoh, S. T., Ogrodzinski, M. P., & Lunt, S. Y. (2020). UDP-glucose 6-dehydrogenase knockout impairs migration and decreases in vivo metastatic ability of breast cancer cells. *Cancer Lett*, *492*, 21-30. doi:10.1016/j.canlet.2020.07.031
- The Cancer Genome Atlas Network, T. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61-70. doi:10.1038/nature11412
- Thiery, J. P. (2002). Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer*, *2*(6), 442-454. doi:10.1038/nrc822
- Thirumurthi, U., Shen, J., Xia, W., LaBaff, A. M., Wei, Y., Li, C.-W., Chang, W.-C., Chen, C.-H., Lin, H.-K., Yu, D., & Hung, M.-C. (2014). MDM2-mediated degradation of SIRT6 phosphorylated by AKT1 promotes tumorigenesis and trastuzumab resistance in breast cancer. *Sci Signal*, *7*(336), ra71. doi:10.1126/scisignal.2005076
- Tian, Z., Zhang, Y., & Lyu, X. (2021). Promoting roles of KLF5 in myocardial infarction in mice involving microRNA-27a suppression and the following GFPT2/TGF- $\beta$ /Smad2/3 axis activation. *Cell Cycle*, 1-20. doi:10.1080/15384101.2021.1907512
- Tokes, T., Tuboly, E., Varga, G., Major, L., Ghyczy, M., Kaszaki, J., & Boros, M. (2015). Protective effects of L-alpha-glycerolphosphorylcholine on ischaemia-reperfusion-induced inflammatory reactions. *Eur J Nutr*, *54*(1), 109-118. doi:10.1007/s00394-014-0691-2
- Tong, M., Che, N., Zhou, L., Luk, S. T., Kau, P. W., Chai, S., Ngan, E. S., Lo, C. M., Man, K., Ding, J., Lee, T. K., & Ma, S. (2018). Efficacy of annexin A3 blockade in sensitizing hepatocellular carcinoma to sorafenib and regorafenib. *J Hepatol*, *69*(4), 826-839. doi:10.1016/j.jhep.2018.05.034
- Tontonoz, P., & Spiegelman, B. M. (2008). Fat and beyond: the diverse biology of PPARgamma. *Annu Rev Biochem*, *77*, 289-312. doi:10.1146/annurev.biochem.77.061307.091829
- Tripathi, S., Chakraborty, P., Levine, H., & Jolly, M. K. (2020). A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. *PLoS Comput Biol*, *16*(2), e1007619. doi:10.1371/journal.pcbi.1007619
- Tsougos, I., Svolos, P., Kousi, E., Fountas, K., Theodorou, K., Fezoulidis, I., & Kapsalaki, E. (2012). Differentiation of glioblastoma multiforme from metastatic brain tumor using proton magnetic resonance spectroscopy, diffusion and

- perfusion metrics at 3 T. *Cancer imaging : the official publication of the International Cancer Imaging Society*, 12(3), 423-436. doi:10.1102/1470-7330.2012.0038
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., & Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, 12(6), 523-526. doi:10.1038/nmeth.3393
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), 731-740. doi:10.1038/nmeth.3901
- Ulgen, E., Ozisik, O., & Sezerman, O. U. (2019). pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. *Frontiers in Genetics*, 10(858). doi:10.3389/fgene.2019.00858
- Ulisse, S., Cinque, B., Silvano, G., Rucci, N., Biordi, L., Cifone, M. G., & D'Armiento, M. (2000). Erk-dependent cytosolic phospholipase A2 activity is induced by CD95 ligand cross-linking in the mouse derived Sertoli cell line TM4 and is required to trigger apoptosis in CD95 bearing cells. *Cell Death Differ*, 7(10), 916-924. doi:10.1038/sj.cdd.4400716
- Van Schaftingen, E., Jett, M. F., Hue, L., & Hers, H. G. (1981). Control of liver 6-phosphofructokinase by fructose 2,6-bisphosphate and other effectors. *Proceedings of the National Academy of Sciences of the United States of America*, 78(6), 3483-3486. doi:10.1073/pnas.78.6.3483
- Vatsyayan, J., Lee, S. J., & Chang, H. Y. (2005). Effects of xenobiotics and peroxisome proliferator-activated receptor-alpha on the human UDPglucose dehydrogenase gene expression. *J Biochem Mol Toxicol*, 19(5), 279-288. doi:10.1002/jbt.20099
- Vecchi, M., Confalonieri, S., Nuciforo, P., Viganò, M. A., Capra, M., Bianchi, M., Nicosia, D., Bianchi, F., Galimberti, V., Viale, G., Palermo, G., Riccardi, A., Campanini, R., Daidone, M. G., Pierotti, M. A., Pece, S., & Di Fiore, P. P. (2008). Breast cancer metastases are molecularly distinct from their primary tumors. *Oncogene*, 27(15), 2148-2158. doi:10.1038/sj.onc.1210858
- Vella, V., De Francesco, E. M., Lappano, R., Muoio, M. G., Manzella, L., Maggiolini, M., & Belfiore, A. (2020). Microenvironmental Determinants of Breast Cancer Metastasis: Focus on the Crucial Interplay Between Estrogen and Insulin/Insulin-Like

- Growth Factor Signaling. *Front Cell Dev Biol*, 8, 608412. doi:10.3389/fcell.2020.608412
- Verbovsek, U., Motaln, H., Rotter, A., Atai, N. A., Gruden, K., Van Noorden, C. J., & Lah, T. T. (2014). Expression analysis of all protease genes reveals cathepsin K to be overexpressed in glioblastoma. *PLoS One*, 9(10), e111819. doi:10.1371/journal.pone.0111819
- Vergara, D., Simeone, P., Latorre, D., Cascione, F., Leporatti, S., Trerotola, M., Giudetti, A. M., Capobianco, L., Lunetti, P., Rizzello, A., Rinaldi, R., Alberti, S., & Maffia, M. (2015). Proteomics analysis of E-cadherin knockdown in epithelial breast cancer cells. *J Biotechnol*, 202, 3-11. doi:10.1016/j.jbiotec.2014.10.034
- Vigetti, D., & Passi, A. (2014). Hyaluronan synthases posttranslational regulation in cancer. *Adv Cancer Res*, 123, 95-119. doi:10.1016/B978-0-12-800092-2.00004-6
- Vitale, D. L., Caon, I., Parnigoni, A., Sevic, I., Spinelli, F. M., Icardi, A., Passi, A., Vigetti, D., & Alaniz, L. (2021). Initial Identification of UDP-Glucose Dehydrogenase as a Prognostic Marker in Breast Cancer Patients, Which Facilitates Epirubicin Resistance and Regulates Hyaluronan Synthesis in MDA-MB-231 Cells. *Biomolecules*, 11(2), 246. Retrieved from <https://www.mdpi.com/2218-273X/11/2/246>  
[https://res.mdpi.com/d\\_attachment/biomolecules/biomolecules-11-00246/article\\_deploy/biomolecules-11-00246.pdf](https://res.mdpi.com/d_attachment/biomolecules/biomolecules-11-00246/article_deploy/biomolecules-11-00246.pdf)
- Voon, D. C., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2017). The EMT spectrum and therapeutic opportunities. *Mol Oncol*, 11(7), 878-891. doi:10.1002/1878-0261.12082
- Wang, H., Yin, Y., Li, W., Zhao, X., Yu, Y., Zhu, J., Qin, Z., Wang, Q., Wang, K., Lu, W., Liu, J., & Huang, L. (2012). Over-expression of PDGFR-beta promotes PDGF-induced proliferation, migration, and angiogenesis of EPCs through PI3K/Akt signaling pathway. *PLoS One*, 7(2), e30503. doi:10.1371/journal.pone.0030503
- Wang, J., Ma, Z., Carr, S. A., Mertins, P., Zhang, H., Zhang, Z., Chan, D. W., Ellis, M. J., Townsend, R. R., Smith, R. D., McDermott, J. E., Chen, X., Paulovich, A. G., Boja, E. S., Mesri, M., Kinsinger, C. R., Rodriguez, H., Rodland, K. D., Liebler, D. C., & Zhang, B. (2017). Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol Cell Proteomics*, 16(1), 121-134. doi:10.1074/mcp.M116.060301



- Wang, L., Zhang, S., & Wang, X. (2020). The Metabolic Mechanisms of Breast Cancer Metastasis. *Front Oncol*, *10*, 602416. doi:10.3389/fonc.2020.602416
- Wang, L., Yang, Y., Ma, H., Xie, Y., Xu, J., Near, D., Wang, H., Garbutt, T., Li, Y., Liu, J., & Qian, L. (2021). Single cell dual-omics reveals the transcriptomic and epigenomic diversity of cardiac non-myocytes. *Cardiovascular Research*. doi:10.1093/cvr/cvab134
- Wang, M., Yan, J., Cao, X., Hua, P., & Li, Z. (2020). Hydrogen sulfide modulates epithelial-mesenchymal transition and angiogenesis in non-small cell lung cancer via HIF-1alpha activation. *Biochem Pharmacol*, *172*, 113775. doi:10.1016/j.bcp.2019.113775
- Wang, P., McKnight, K. D., Wong, D. J., Rodriguez, R. T., Sugiyama, T., Gu, X., Ghodasara, A., Qu, K., Chang, H. Y., & Kim, S. K. (2012). A molecular signature for purified definitive endoderm guides differentiation and isolation of endoderm from mouse and human embryonic stem cells. *Stem Cells Dev*, *21*(12), 2273-2287. doi:10.1089/scd.2011.0416
- Wang, Q., Karvelsson, S. T., Johannsson, F., Vilhjalmsson, A. I., Hagen, L., de Miranda Fonseca, D., Sharma, A., Slupphaug, G., & Rolfsson, O. (2021). UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines. *Mol Oncol*. doi:10.1002/1878-0261.13172
- Wang, Q., Karvelsson, S. T., Kotronoulas, A., Gudjonsson, T., Halldorsson, S., & Rolfsson, O. (2021). Glutamine-Fructose-6-Phosphate Transaminase 2 (GFPT2) Is Upregulated in Breast Epithelial-Mesenchymal Transition and Responds to Oxidative Stress. *Mol Cell Proteomics*, *21*(2), 100185. doi:10.1016/j.mcpro.2021.100185
- Wang, T. P., Pan, Y. R., Fu, C. Y., & Chang, H. Y. (2010). Down-regulation of UDP-glucose dehydrogenase affects glycosaminoglycans synthesis and motility in HCT-8 colorectal carcinoma cells. *Exp Cell Res*, *316*(17), 2893-2902. doi:10.1016/j.yexcr.2010.07.017
- Wang, X., Liu, R., Zhu, W., Chu, H., Yu, H., Wei, P., Wu, X., Zhu, H., Gao, H., Liang, J., Li, G., & Yang, W. (2019). UDP-glucose accelerates SNAIL mRNA decay and impairs lung cancer metastasis. *Nature*, *571*(7763), 127-131. doi:10.1038/s41586-019-1340-y

- Wang, Z., Ahmad, A., Li, Y., Kong, D., Azmi, A. S., Banerjee, S., & Sarkar, F. H. (2010). Emerging roles of PDGF-D signaling pathway in tumor development and progression. *Biochim Biophys Acta*, 1806(1), 122-130. doi:10.1016/j.bbcan.2010.04.003
- Warburg, O., Wind, F., & Negelein, E. (1927). THE METABOLISM OF TUMORS IN THE BODY. *The Journal of general physiology*, 8(6), 519-530. doi:10.1085/jgp.8.6.519
- Weber, R., & Birsoy, K. (2019). The Transsulfuration Pathway Makes, the Tumor Takes. *Cell Metab*, 30(5), 845-846. doi:10.1016/j.cmet.2019.10.009
- Wellen, K. E., Lu, C., Mancuso, A., Lemons, J. M., Ryczko, M., Dennis, J. W., Rabinowitz, J. D., Collier, H. A., & Thompson, C. B. (2010). The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes Dev*, 24(24), 2784-2799. doi:10.1101/gad.1985910
- Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., zur Hausen, A., Brunton, V. G., Morton, J., Sansom, O., Schüler, J., Stemmler, M. P., Herzberger, C., Hopt, U., Keck, T., Brabletz, S., & Brabletz, T. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol*, 11(12), 1487-1495. doi:10.1038/ncb1998
- Wen, Y., Li, J., Wang, L., Tie, K., Magdalou, J., Chen, L., & Wang, H. (2014). UDP-glucose dehydrogenase modulates proteoglycan synthesis in articular chondrocytes: its possible involvement and regulation in osteoarthritis. *Arthritis Res Ther*, 16(6), 484. doi:10.1186/s13075-014-0484-2
- Whitehead, R. H., Bertoncello, I., Webber, L. M., & Pedersen, J. S. (1983). A new human breast carcinoma cell line (PMC42) with stem cell characteristics. I. Morphologic characterization. *J Natl Cancer Inst*, 70(4), 649-661.
- Williams, E. D., Gao, D., Redfern, A., & Thompson, E. W. (2019). Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nat Rev Cancer*, 19(12), 716-732. doi:10.1038/s41568-019-0213-x
- Wilson, M. M., Weinberg, R. A., Lees, J. A., & Guen, V. J. (2020). Emerging Mechanisms by which EMT Programs Control Stemness. *Trends Cancer*, 6(9), 775-780. doi:10.1016/j.trecan.2020.03.011

- Wirsik, N. M., Ehlers, J., Mäder, L., Ilina, E. I., Blank, A. E., Grote, A., Feuerhake, F., Baumgarten, P., Devraj, K., Harter, P. N., Mittelbronn, M., & Naumann, U. (2021). TGF- $\beta$  activates pericytes via induction of the epithelial-to-mesenchymal transition protein SLUG in glioblastoma. *Neuropathol Appl Neurobiol*, *47*(6), 768-780. doi:10.1111/nan.12714
- Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., & Scalbert, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*, *46*(D1), D608-d617. doi:10.1093/nar/gkx1089
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359-362. doi:10.1038/nmeth.1322
- Wong, I. Y., Javaid, S., Wong, E. A., Perk, S., Haber, D. A., Toner, M., & Irimia, D. (2014). Collective and individual migration following the epithelial-mesenchymal transition. *Nat Mater*, *13*(11), 1063-1071. doi:10.1038/nmat4062
- Wu, Q., Wang, R., Yang, Q., Hou, X., Chen, S., Hou, Y., Chen, C., Yang, Y., Miele, L., Sarkar, F., Chen, Y., & Wang, Z. (2013). Chemoresistance to gemcitabine in hepatoma cells induces epithelial-mesenchymal transition and involves activation of PDGF-D pathway. *Oncotarget*, *4*(11). Retrieved from <https://www.oncotarget.com/article/1471/text/>
- Wu, X., Li, X., Fu, Q., Cao, Q., Chen, X., Wang, M., Yu, J., Long, J., Yao, J., Liu, H., Wang, D., Liao, R., & Dong, C. (2017). AKR1B1 promotes basal-like breast cancer progression by a positive feedback loop that activates the EMT program. *J Exp Med*, *214*(4), 1065-1079. doi:10.1084/jem.20160903
- Wynn, M. L., Yates, J. A., Evans, C. R., Van Wassenhove, L. D., Wu, Z. F., Bridges, S., Bao, L., Fournier, C., Ashrafzadeh, S., Merrins, M. J., Satin, L. S., Schnell, S., Burant, C. F., & Merajver, S. D. (2016). RhoC GTPase Is a Potent Regulator of Glutamine Metabolism and N-Acetylaspartate Production in Inflammatory Breast Cancer Cells. *J Biol Chem*, *291*(26), 13715-13729. doi:10.1074/jbc.M115.703959
- Xu, F., Zhang, J., Hu, G., Liu, L., & Liang, W. (2017). Hypoxia and TGF- $\beta$ 1 induced PLOD2 expression improve the migration and invasion of cervical cancer cells by promoting epithelial-to-mesenchymal transition (EMT) and focal adhesion formation.

*Cancer cell international*, 17, 54-54. doi:10.1186/s12935-017-0420-z

- Xu, X., Qiao, D., Mann, M., Garofalo, R. P., & Brasier, A. R. (2020). Respiratory Syncytial Virus Infection Induces Chromatin Remodeling to Activate Growth Factor and Extracellular Matrix Secretion Pathways. *Viruses*, 12(8). doi:10.3390/v12080804
- Xu, X. M., Zhou, Y. Q., & Wang, M. H. (2005). Mechanisms of cytoplasmic {beta}-catenin accumulation and its involvement in tumorigenic activities mediated by oncogenic splicing variant of the receptor originated from Nantes tyrosine kinase. *J Biol Chem*, 280(26), 25087-25094. doi:10.1074/jbc.M414699200
- Yamagami, Y., Kawami, M., Ojima, T., Futatsugi, S., Yumoto, R., & Takano, M. (2020). Role of plasminogen activator inhibitor-1 in methotrexate-induced epithelial-mesenchymal transition in alveolar epithelial A549 cells. *Biochem Biophys Res Commun*, 525(3), 543-548. doi:10.1016/j.bbrc.2020.02.131
- Yan, C., Zhang, X., Miao, J., Yuan, H., Liu, E., Liang, T., & Li, Q. (2020). Farrerol Directly Targets GSK-3beta to Activate Nrf2-ARE Pathway and Protect EA.hy926 Cells against Oxidative Stress-Induced Injuries. *Oxid Med Cell Longev*, 2020, 5967434. doi:10.1155/2020/5967434
- Yan, G. R., Xu, S. H., Tan, Z. L., Liu, L., & He, Q. Y. (2011). Global identification of miR-373-regulated genes in breast cancer by quantitative proteomics. *Proteomics*, 11(5), 912-920. doi:10.1002/pmic.201000539
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., & Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*, 117(7), 927-939. doi:10.1016/j.cell.2004.06.006
- Yang, J., Kong, P., Yang, J., Jia, Z., Hu, X., Wang, Z., Cui, H., Bi, Y., Qian, Y., Li, H., Wang, F., Yang, B., Yan, T., Ma, Y., Zhang, L., Cheng, C., Song, B., Li, Y., Xu, E., Liu, H., Gao, W., Wang, J., Liu, Y., Zhai, Y., Chang, L., Wang, Y., Zhang, Y., Shi, R., Liu, J., Wang, Q., Cheng, X., & Cui, Y. (2018). High TSTA3 Expression as a Candidate Biomarker for Poor Prognosis of Patients With ESCC. *Technol Cancer Res Treat*, 17, 1533033818781405. doi:10.1177/1533033818781405
- Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar,

- S., Derynck, R., Ford, H. L., Fuxe, J., Garcia de Herreros, A., Goodall, G. J., Hadjantonakis, A. K., Huang, R. J. Y., Kalchauer, C., Kalluri, R., Kang, Y., Khew-Goodall, Y., Levine, H., Liu, J., Longmore, G. D., Mani, S. A., Massague, J., Mayor, R., McClay, D., Mostov, K. E., Newgreen, D. F., Nieto, M. A., Puisieux, A., Runyan, R., Savagner, P., Stanger, B., Stemmler, M. P., Takahashi, Y., Takeichi, M., Theveneau, E., Thiery, J. P., Thompson, E. W., Weinberg, R. A., Williams, E. D., Xing, J., Zhou, B. P., Sheng, G., & Association, E. M. T. I. (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. doi:10.1038/s41580-020-0237-9
- Yang, M., & Vousden, K. H. (2016). Serine and one-carbon metabolism in cancer. *Nat Rev Cancer*, 16(10), 650-662. doi:10.1038/nrc.2016.81
- Yang, X. G., Zhu, L. C., Wang, Y. J., Li, Y. Y., & Wang, D. (2019). Current Advance of Therapeutic Agents in Clinical Trials Potentially Targeting Tumor Plasticity. *Front Oncol*, 9, 887. doi:10.3389/fonc.2019.00887
- Ye, X., Wei, X., Liao, J., Chen, P., Li, X., Chen, Y., Yang, Y., Zhao, Q., Sun, H., Pan, L., Chen, G., He, X., Lyu, J., & Fang, H. (2021). 4-Hydroxyphenylpyruvate Dioxygenase-Like Protein Promotes Pancreatic Cancer Cell Progression and Is Associated With Glutamine-Mediated Redox Balance. *Frontiers in Oncology*, 10(3074). doi:10.3389/fonc.2020.617190
- Yi, W., Clark, P. M., Mason, D. E., Keenan, M. C., Hill, C., Goddard, W. A., 3rd, Peters, E. C., Driggers, E. M., & Hsieh-Wilson, L. C. (2012). Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science*, 337(6097), 975-980. doi:10.1126/science.1222278
- Yogev, O., Yogev, O., Singer, E., Shaulian, E., Goldberg, M., Fox, T. D., & Pines, O. (2010). Fumarase: A Mitochondrial Metabolic Enzyme and a Cytosolic/Nuclear Component of the DNA Damage Response. *PLOS Biology*, 8(3), e1000328. doi:10.1371/journal.pbio.1000328
- Yook, J. I., Li, X. Y., Ota, I., Hu, C., Kim, H. S., Kim, N. H., Cha, S. Y., Ryu, J. K., Choi, Y. J., Kim, J., Fearon, E. R., & Weiss, S. J. (2006). A Wnt-Axin2-GSK3beta cascade regulates Snail activity in breast cancer cells. *Nat Cell Biol*, 8(12), 1398-1406. doi:10.1038/ncb1508
- Young, R. K., Cailleau, R. M., Mackay, B., & Reeves, W. J. (1974). Establishment of epithelial cell line MDA-MB-157 from

- metastatic pleural effusion of human breast carcinoma. *In Vitro*, 9(4), 239-245. doi:10.1007/BF02616069
- Yousefnia, S., Momenzadeh, S., Seyed Forootan, F., Ghaedi, K., & Nasr Esfahani, M. H. (2018). The influence of peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ) ligands on cancer cell tumorigenicity. *Gene*, 649, 14-22. doi:10.1016/j.gene.2018.01.018
- Yu, H., Wang, L., Lv, L., Ma, C., Du, B., Lu, T., Jin, C., Yan, H., Yang, Y., Li, W., Ruan, Y., Zhang, H., Zhang, H., Mi, W., Mowry, B., Ma, W., Li, K., Zhang, D., & Yue, W. (2016). Genome-Wide Association Study Suggested the PTPRD Polymorphisms Were Associated With Weight Gain Effects of Atypical Antipsychotic Medications. *Schizophr Bull*, 42(3), 814-823. doi:10.1093/schbul/sbv179
- Yu, H., Guo, W., Liu, Y., & Wang, Y. (2021). Immune Characteristics Analysis and Transcriptional Regulation Prediction Based on Gene Signatures of Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis*, 16, 3027-3039. doi:10.2147/copd.S325328
- Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., Isakoff, S. J., Ciciliano, J. C., Wells, M. N., Shah, A. M., Concannon, K. F., Donaldson, M. C., Sequist, L. V., Brachtel, E., Sgroi, D., Baselga, J., Ramaswamy, S., Toner, M., Haber, D. A., & Maheswaran, S. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*, 339(6119), 580-584. doi:10.1126/science.1228522
- Yu, Y., Luo, W., Yang, Z. J., Chi, J. R., Li, Y. R., Ding, Y., Ge, J., Wang, X., & Cao, X. C. (2018). miR-190 suppresses breast cancer metastasis by regulation of TGF-beta-induced epithelial-mesenchymal transition. *Mol Cancer*, 17(1), 70. doi:10.1186/s12943-018-0818-9
- Zaidi, N., Swinnen, J. V., & Smans, K. (2012). ATP-citrate lyase: a key player in cancer metabolism. *Cancer Res*, 72(15), 3709-3714. doi:10.1158/0008-5472.CAN-11-4112
- Zand, B., Previs, R. A., Zacharias, N. M., Rupaimoole, R., Mitamura, T., Nagaraja, A. S., Guindani, M., Dalton, H. J., Yang, L., Baddour, J., Achreja, A., Hu, W., Pecot, C. V., Ivan, C., Wu, S. Y., McCullough, C. R., Gharpure, K. M., Shoshan, E., Pradeep, S., Mangala, L. S., Rodriguez-Aguayo, C., Wang, Y., Nick, A. M., Davies, M. A., Armaiz-Pena, G., Liu, J., Lutgendorf, S. K., Baggerly, K. A., Eli, M. B., Lopez-Berestein, G., Nagrath, D., Bhattacharya, P. K., & Sood, A. K. (2016). Role of Increased

- n-acetylaspartate Levels in Cancer. *J Natl Cancer Inst*, 108(6), djv426. doi:10.1093/jnci/djv426
- Zeng, B., Ge, C., Li, R., Zhang, Z., Fu, Q., Li, Z., Lin, Z., Liu, L., Xue, Y., Xu, Y., He, J., Guo, H., Li, C., Huang, W., Song, X., & Huang, Y. (2020). Knockdown of microsomal glutathione S-transferase 1 inhibits lung adenocarcinoma cell proliferation and induces apoptosis. *Biomed Pharmacother*, 121, 109562. doi:10.1016/j.biopha.2019.109562
- Zhang, D., Xu, X., & Ye, Q. (2021). Metabolism and immunity in breast cancer. *Front Med*, 15(2), 178-207. doi:10.1007/s11684-020-0793-6
- Zhang, H., Jia, Y., Cooper, J. J., Hale, T., Zhang, Z., & Elbein, S. C. (2004). Common variants in glutamine:fructose-6-phosphate amidotransferase 2 (GFPT2) gene are associated with type 2 diabetes, diabetic nephropathy, and increased GFPT2 mRNA levels. *J Clin Endocrinol Metab*, 89(2), 748-755. doi:10.1210/jc.2003-031286
- Zhang, J., Jia, L., Liu, T., Yip, Y. L., Tang, W. C., Lin, W., Deng, W., Lo, K. W., You, C., Lung, M. L., Lung, H. L., Cheung, A. L., Tsao, S. W., & Tsang, C. M. (2019). mTORC2-mediated PDHE1 $\alpha$  nuclear translocation links EBV-LMP1 reprogrammed glucose metabolism to cancer metastasis in nasopharyngeal carcinoma. *Oncogene*, 38(24), 4669-4684. doi:10.1038/s41388-019-0749-y
- Zhang, L., Sun, W., Ren, W., Zhang, J., & Xu, G. (2021). Predicting Panel of Metabolism and Immune-Related Genes for the Prognosis of Human Ovarian Cancer. *Front Cell Dev Biol*, 9, 690542. doi:10.3389/fcell.2021.690542
- Zhang, W., Bouchard, G., Yu, A., Shafiq, M., Jamali, M., Shrager, J. B., Ayers, K., Bakr, S., Gentles, A. J., Diehn, M., Quon, A., West, R. B., Nair, V., van de Rijn, M., Napel, S., & Plevritis, S. K. (2018). GFPT2-Expressing Cancer-Associated Fibroblasts Mediate Metabolic Reprogramming in Human Lung Adenocarcinoma. *Cancer Res*, 78(13), 3445-3457. doi:10.1158/0008-5472.Can-17-2928
- Zhang, Z. G., Zhang, H. S., Sun, H. L., Liu, H. Y., Liu, M. Y., & Zhou, Z. (2019). KDM5B promotes breast cancer cell proliferation and migration via AMPK-mediated lipid metabolism reprogramming. *Exp Cell Res*, 379(2), 182-190. doi:10.1016/j.yexcr.2019.04.006

- Zhao, M., Kong, L., Liu, Y., & Qu, H. (2015). dbEMT: an epithelial-mesenchymal transition associated gene resource. *Scientific Reports*, *5*, 11459. doi:10.1038/srep11459
- Zhao, M., Liu, Y., Zheng, C., & Qu, H. (2019). dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. *J Genet Genomics*, *46*(12), 595-597. doi:10.1016/j.jgg.2019.11.010
- Zhao, Y., Chahar, H. S., Komaravelli, N., Dossumentkova, A., & Casola, A. (2019). Human metapneumovirus infection of airway epithelial cells is associated with changes in core metabolic pathways. *Virology*, *531*, 183-191. doi:10.1016/j.virol.2019.03.011
- Zhao, Z., Wu, M. S., Zou, C., Tang, Q., Lu, J., Liu, D., Wu, Y., Yin, J., Xie, X., Shen, J., Kang, T., & Wang, J. (2014). Downregulation of MCT1 inhibits tumor growth, metastasis and enhances chemotherapeutic efficacy in osteosarcoma through regulation of the NF- $\kappa$ B pathway. *Cancer Lett*, *342*(1), 150-158. doi:10.1016/j.canlet.2013.08.042
- Zhou, L., Luo, M., Cheng, L. J., Li, R. N., Liu, B., & Linghu, H. (2019). Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) promotes the EMT of serous ovarian cancer by activating the hexosamine biosynthetic pathway to increase the nuclear location of beta-catenin. *Pathol Res Pract*, *215*(12), 152681. doi:10.1016/j.prp.2019.152681
- Zhu, J., Berisa, M., Schworer, S., Qin, W., Cross, J. R., & Thompson, C. B. (2019). Transsulfuration Activity Can Support Cell Growth upon Extracellular Cysteine Limitation. *Cell Metab*, *30*(5), 865-876 e865. doi:10.1016/j.cmet.2019.09.009
- Zhu, M., Lu, T., Jia, Y., Luo, X., Gopal, P., Li, L., Odewole, M., Renteria, V., Singal, A. G., Jang, Y., Ge, K., Wang, S. C., Sorouri, M., Parekh, J. R., MacConmara, M. P., Yopp, A. C., Wang, T., & Zhu, H. (2019). Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell*, *177*(3), 608-621.e612. doi:10.1016/j.cell.2019.03.026
- Zimmer, B. M., Howell, M. E., Wei, Q., Ma, L., Romsdahl, T., Loughman, E. G., Markham, J. E., Seravalli, J., Barycki, J. J., & Simpson, M. A. (2016). Loss of exogenous androgen dependence by prostate tumor cells is associated with elevated glucuronidation potential. *Horm Cancer*, *7*(4), 260-271. doi:10.1007/s12672-016-0268-z




Zitzler, J., Link, D., Schafer, R., Liebetrau, W., Kazinski, M., Bonin-Debs, A., Behl, C., Buckel, P., & Brinkmann, U. (2004). High-throughput functional genomics identifies genes that ameliorate toxicity due to oxidative stress in neuronal HT-22 cells: GFPT2 protects cells against peroxide. *Mol Cell Proteomics*, 3(8), 834-840. doi:10.1074/mcp.M400054-MCP200



# Paper I



## UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines

Qiong Wang<sup>1</sup>, Sigurdur Trausti Karvelsson<sup>1</sup>, Freyr Johannsson<sup>1</sup>, Arnar Ingi Vilhjalmsón<sup>1</sup>, Lars Hagen<sup>2,3,4</sup>, Davi de Miranda Fonseca<sup>2,3,4</sup>, Animesh Sharma<sup>2,3,4</sup>, Geir Slupphaug<sup>2,3,4</sup> and Ottar Rolfsson<sup>1</sup> 

1 Center for Systems Biology, Biomedical Center, Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland  
 2 Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, NTNU, Trondheim, Norway  
 3 Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim, Norway  
 4 PROMEC Core Facility for Proteomics and Modomics, Norwegian University of Science and Technology, NTNU, and the Central Norway Regional Health Authority Norway, Trondheim, Norway

### Keywords

acetylaspartate; breast cancer; EMT; glycerophosphocholine; PDGFRB; UGDH

### Correspondence

O. Rolfsson, Center for Systems Biology, Biomedical Center, Faculty of Medicine, School of Health Sciences, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland  
 Tel: +354 5255854  
 E-mail: ottarr@hi.is

(Received 6 August 2021, revised 4 November 2021, accepted 21 December 2021)

doi:10.1002/1878-0261.13172

Metabolic rewiring is one of the indispensable drivers of epithelial–mesenchymal transition (EMT) involved in breast cancer metastasis. In this study, we explored the metabolic changes during spontaneous EMT in three separately established breast EMT cell models using a proteomic approach supported by metabolomic analysis. We identified common proteomic changes, including the expression of CDH1, CDH2, VIM, LGALS1, SERPINE1, PKP3, ATP2A2, JUP, MTCH2, RPL26L1 and PLOD2. Consistently altered metabolic enzymes included the following: FDFT1, SORD, TSTA3 and UDP-glucose dehydrogenase (UGDH). Of these, UGDH was most prominently altered and has previously been associated with breast cancer patient survival. siRNA-mediated knock-down of *UGDH* resulted in delayed cell proliferation and dampened invasive potential of mesenchymal cells and downregulated expression of the EMT transcription factor SNAIL. Metabolomic analysis revealed that siRNA-mediated knock-down of *UGDH* decreased intracellular glycerophosphocholine (GPC), whereas levels of acetylaspartate (NAA) increased. Finally, our data suggested that platelet-derived growth factor receptor beta (PDGFRB) signalling was activated in mesenchymal cells. siRNA-mediated knock-down of *PDGFRB* downregulated *UGDH* expression, potentially via NFκB-p65. Our results support an unexplored relationship between UGDH and GPC, both of which have previously been independently associated with breast cancer progression.

### Abbreviations

AGC, automatic gain control; ATP2A2, sarcoplasmic/endoplasmic reticulum calcium ATPase 2; CD44, CD44 antigen; CDH1, E-cadherin; CDH2, N-cadherin; EGFR, epidermal growth factor receptor; EMT, epithelial–mesenchymal transition; FDFT1, squalene synthase; FDR, false discovery rate; FLNA, filamin-A; FSCN1, fascin; GEMs, genome-scale metabolic network reconstructions; GPC, glycerophosphocholine; HBP, hexosamine biosynthesis pathway; JUP, junction plakoglobin; LGALS1, galectin-1; LMNB1, Lamin-B1; MSN, moesin; MTCH2, mitochondrial carrier homolog 2; NAA, acetylaspartate; NDRG1, N-myc downstream regulated 1; PDGFRB, platelet-derived growth factor receptor beta; pERK, extracellular signal-regulated kinase (phosphorylated); PKP3, plakophilin-3; PLA2G15, phospholipase A2 group XV; PLOD2, procollagen-lysine,2-oxoglutarate; PPARγ, peroxisome proliferator-activated receptor γ; PSM, peptide-spectrum matches; RELA (NFκB-p65), nuclear factor NF-kappa-B p65 subunit; RPL26L1, 60S ribosomal protein L26-like 1; S100A2, S100 calcium-binding protein A2; SERPINE1, plasminogen activator inhibitor 1; SNAIL, Snail family transcriptional repressor 1; SORD, sorbitol dehydrogenase; TLN1, Talin-1; TSTA3, GDP-L-fucose synthase; UDP-Glc, UDP-glucose; UDP-GlcA, UDP-glucuronate; UGDH, UDP-glucose dehydrogenase; VIM, vimentin.

## 1. Introduction

Epithelial-mesenchymal transition (EMT) is a core developmental process that allows a polarized epithelial cell to assume mesenchymal phenotypes through a series of morphological, molecular, regulatory and functional changes [1]. EMT is part of normal tissue development, organ/tissue fibrosis, wound healing and cancer malignancy. Partial activation of EMT drives tumour metastasis and dissemination to distant organs [2,3].

The biological plasticity and molecular heterogeneity of the EMT programme indicate that EMT is context-specific, which has resulted in discrepancies in its description in the literature [4]. One factor causing EMT heterogeneity in cell line models is the EMT induction method. Growth factors, transcription factors and microRNAs can be manipulated in cells to trigger EMT, such as TGF- $\beta$ , EGF, Snail1/2, ZEB1 and Twist [3,5–8]. Genetic manipulation may lead to effects that diminish the flexibility and plasticity of the EMT programme. In the present study, we used three breast EMT cell models related to basal mammary cells to investigate common effects of the spontaneous EMT programme, that is, D492 EMT cell lines [9,10], HMLE EMT cell lines [7,11,12] and PMC42 EMT cell lines [13–15]. Each EMT cell model includes a breast epithelial/mesenchymal cell line pair generated with spontaneous induction methods without overexpressing specific EMT inducers. These EMT models may reside at different positions in the EMT spectrum [1] but represent typical EMT progress in human breast gland development and have contributed significantly to the understanding of the molecular regulatory machinery in EMT [16–18].

Metabolic reprogramming is an indispensable driver of EMT in cancer [19]. A better understanding of the metabolism of EMT may facilitate the development of new therapeutics for breast cancer treatment. Glucose metabolism, lipid metabolism, an acidic microenvironment, nucleotide metabolism and amino acid metabolism have been related to EMT in cancer malignancy [20,21]. In our previous studies on EMT-related metabolic dysregulation limited to the D492 EMT model, we observed different preferences for reductive/oxidative carboxylation, glycolysis and amino acid anaplerosis along with an altered lipid profile and shifted glutathione homeostasis [22–24]. To move towards a system's understanding of how metabolism is influenced following EMT, we compared the metabolic phenotypes of EMT within genome-scale metabolic network reconstructions (GEMs) that allow the integrated analysis of gene expression, proteomic and metabolomic data. These analyses revealed increased

dependency on argininosuccinate lyase (ASL) and enhanced activities of the pentose phosphate pathway, hexosamine biosynthesis and one-carbon metabolism post-EMT in the D492 EMT model [23,25]. More recently, we confirmed that metabolic flux through the hexosamine biosynthesis pathway (HBP) increases significantly in mesenchymal cells and that glutamine-fructose-6-phosphate transaminase 2 (GFPT2) in the HBP is associated with breast cancer malignancy [26].

In this study, we further explored the metabolic changes in EMT using shotgun proteomics and expanded our analysis to include three breast cell models descriptive of spontaneous EMT (Fig. 1 and Fig. S1). Several metabolic enzymes were commonly changed after EMT, with UDP-glucose dehydrogenase (UGDH) being most altered. UGDH catalyses conversion of UDP-glucose (UDP-Glc) to UDP-glucuronate (UDP-GlcA), both of which are essential metabolites with diverse cellular functions [27,28]. UGDH is involved in a variety of regulatory events. SP1, TGF- $\beta$ , Slit2, p38<sup>MAPK</sup> and PI3K/Akt regulate UGDH expression, which in turn influences the downstream targets ERK/MAPK, PPAR $\gamma$  and SNAI1 [27–34]. Several studies have recently reported that UGDH is involved in tumour growth, metastasis and patient survival [27,32,34–37]. To understand the roles of UGDH in EMT in the breast gland, we knocked down *UGDH* in breast mesenchymal cells with siRNAs and studied effects on cell function and metabolism. Importantly, the three EMT models studied are noncarcinogenic. To account for UGDH in cancer progression and oncogenesis, we compared results from the EMT models to UGDH functions in the tumorigenic breast mesenchymal cell lines D492HER2 and MDA-MB-231. These investigations suggest that the tumour promoting effects of UGDH may in part be attributed to changes in choline metabolism.

## 2. Materials and methods

### 2.1. Cell culture

D492 was isolated from primary cultures of reduction mammaplasties with immortalization [9]. D492M was generated via 3D coculture of the D492 cells with endothelial cells to induce EMT [10]. D492HER2 was established by the overexpression of HER2 receptors on D492 [38]. The D492 cell lines (D492, D492M and D492HER2) were cultured in serum-free H14 medium (DMEM/F-12 without glutamine; Thermo Fisher Scientific (TFS), Waltham, MA, USA; 21331020) supplemented with 250 ng·mL<sup>-1</sup> insulin (Merck, Kenilworth, NJ,

	D492-D492M	HMLE-HMLEM	PMC42LA-PMC42ET
<b>Tissue</b>	Breast reduction mammoplasty	Breast reduction mammoplasty	Breast; derived from metastatic site: pleural effusion
<b>Immortalization</b>	HPV-16 E6/E7 gene transduction	hTERT and SV-40 Large T antigen expression	N/A
<b>Cell type</b>	Basal-like; with stem cell properties	Luminal and basal epithelial cells	Express secretory and myoepithelial markers
<b>Induce EMT</b>	Spontaneous EMT; Co-culture with endothelial cells	Spontaneous EMT; Trypsinization	Spontaneous MET; Treatment with a combination of hormones to induce MET
<b>Tumorigenicity</b>	Non-tumorigenic	Non-tumorigenic	Non-tumorigenic

**Fig. 1.** Summary of the three breast EMT cell models. Comparison of the three breast EMT cell models with respect to tissue origin, immortalization methods, cell markers, EMT induction methods and tumorigenicity.

USA; I6634), 10  $\mu\text{g}\cdot\text{mL}^{-1}$  transferrin (Merck; T2252), 10  $\text{ng}\cdot\text{mL}^{-1}$  EGF (PeproTech, Cranbury, NJ, USA; AF-100-15), 2.6  $\text{ng}\cdot\text{mL}^{-1}$  Na-selenite (BD Biosciences, San Jose, CA, USA; 354201),  $10^{-10}$  M estradiol (Sigma, St. Louis, MO, USA; E2758),  $1.4 \times 10^{-6}$  M hydrocortisone (Sigma; H0888), 0.15 IU prolactin (PeproTech; 100-07), 100 IU penicillin and 0.1  $\text{mg}\cdot\text{mL}^{-1}$  streptomycin (Gibco™, TFS; 15140122) and 2 mM glutamine (TFS; 25030024). The passage numbers for both D492 and D492M were from 31, while D492HER2 was cultured from passage 65. HMLE was isolated from reduction mammoplasties [12], while HMLEM was generated from HMLE via differential trypsinization [7]. The HMLE cell lines (HMLE with passage number from 16 and HMLEM with passage number from 28) were cultured in serum-free DMEM/F-12 medium supplemented with 10  $\mu\text{g}\cdot\text{mL}^{-1}$  insulin, 10  $\text{ng}\cdot\text{mL}^{-1}$  EGF,  $1.4 \times 10^{-6}$  M hydrocortisone, 100 IU penicillin and 0.1  $\text{mg}\cdot\text{mL}^{-1}$  streptomycin and 2 mM glutamine. PMC42ET was originally established from pleural effusion from the metastatic site in a breast cancer patient [15], and PMC42LA was then generated via mesenchymal-epithelial transition by hormone treatments [13]. The PMC42 cell lines (PMC42LA and PMC42ET with passage numbers from 9) were cultured in RPMI 1640 Medium (TFS; 52400025) supplemented with 10% FBS (Gibco™ 10270106) and 100 IU penicillin and 0.1  $\text{mg}\cdot\text{mL}^{-1}$  streptomycin. The antibiotics were excluded in the medium for the transient knock-down experiments. In the SILAC phosphoproteomic

experiment, DMEM/F-12 was replaced by 'DMEM:F-12 for SILAC' (TFS; 88370) with light- (l-arginine, l-lysine), medium- [l-arginine- $^{13}\text{C}_6$  hydrochloride (Arg + 6 Da), l-lysine-4,4,5,5-d4 hydrochloride (Lys + 4 Da)] or heavy- (l-arginine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$  hydrochloride (Arg + 10 Da), l-lysine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_2$  hydrochloride (Lys + 8 Da)] labelled arginine or lysine (Cambridge Isotope Laboratories, Tewksbury, MA, USA). In the invasion assay, H14 was supplemented with 10% FBS in the lower chamber of the Transwell. The MDA-MB-231 cells (passage number 26) were cultured in RPMI 1640 supplemented with 10% FBS and 100 IU penicillin and 0.1  $\text{mg}\cdot\text{mL}^{-1}$  streptomycin. All cell lines were cultured at 37 °C with 5%  $\text{CO}_2$  for routine maintenance, and cells were routinely checked for mycoplasma contamination. All cell lines used in this study were kindly provided by the Stem Cell Research Unit, Biomedical Center, University of Iceland.

## 2.2. LFQ proteomics and SILAC phosphoproteomic analysis

The proteomic experimental set-up was illustrated in Fig. S1A.

### 2.2.1. LFQ protein and peptide sample preparation

Cells were lysed with 4% sodium dodecyl sulfate (SDS; MP Biomedicals™, Irvine, CA, USA) in

100 mM Tris (Sigma) and kept on ice for 10 min and then transferred to 1.5-mL Eppendorf tubes. After five freeze/thaw ( $-80^{\circ}\text{C}/\text{room temperature}$ ) cycles, the samples were centrifuged at 20 817 *g* for 20 min at  $4^{\circ}\text{C}$ . The supernatants were collected and aliquoted in new tubes and stored at  $-80^{\circ}\text{C}$ . Total protein was quantified with the Pierce<sup>TM</sup> BCA protein assay (TFS). A volume containing 12–15  $\mu\text{g}$  total protein was precipitated by chloroform/methanol precipitation and reconstituted in 50 mM ammonium bicarbonate. The protein sample was reduced with 1 M dithiothreitol (DTT) for 20 min at  $70^{\circ}\text{C}$  and then alkylated by 200 mM iodoacetamide (IAM) at room temperature in the dark for 30 min, followed by quenching the extra IAM with 1 M DTT for 20 min at room temperature in the dark. Samples were digested overnight with 1.5  $\mu\text{g}$  trypsin at  $37^{\circ}\text{C}$ . Tryptic peptides were desalted using C-18 StageTips as described [39], after which peptides were dried in a SpeedVac centrifuge and resuspended in 0.1% formic acid.

### 2.2.2. LFQ LC-MS/MS analysis

Peptides were analysed on an LC-MS/MS platform consisting of an Easy-nLC 1200 UHPLC system (TFS) interfaced with a QExactive HF Orbitrap Mass Spectrometer (TFS) via a nanospray ESI ion source (TFS). Peptides were injected into a C-18 trap column (Acclaim PepMap100, 75  $\mu\text{m}$  i. d.  $\times$  2 cm, C-18, 3  $\mu\text{m}$ , 100  $\text{\AA}$ ; TFS) and further separated on a C-18 analytical column (Acclaim PepMap100, 75  $\mu\text{m}$  i. d.  $\times$  50 cm, C-18, 2  $\mu\text{m}$ , 100  $\text{\AA}$ ; TFS) using a multi-step gradient with buffer A (0.1% formic acid) and buffer B (80%  $\text{CH}_3\text{CN}$ , 0.1% formic acid): from 2% to 10% buffer B in 10 min, 10% to 50% buffer B in 130 min, 50% to 100% buffer B in 20 min and 20 min with 100% buffer B. The HPLC was re-equilibrated with 2% buffer B before the next injection. The flow rate was 250  $\text{nL}\cdot\text{min}^{-1}$ . Peptides eluted were analysed on QExactive HF mass spectrometer (TFS) operating in positive ion- and data-dependent acquisition mode using the following parameters: electrospray voltage 1.9 kV, HCD fragmentation with normalized collision energy 29, automatic gain control (AGC) target value of  $3 \times 10^6$  for Orbitrap MS and  $1 \times 10^5$  for MS/MS scans. Each MS scan ( $m/z$  350–1650) was acquired at a resolution of 120 000 FWHM, followed by 15 MS/MS scans triggered for AGC targets above  $2 \times 10^3$ , at a maximum ion injection time of 100 ms for MS and 100 ms for MS/MS scans. The proteomic method has been described previously [40].

### 2.2.3. LFQ protein and peptide identification and quantification

Proteins were identified and quantified by processing MS data using Thermo Scientific<sup>TM</sup> PROTEOME DISCOVERER<sup>TM</sup> (PD, version 2.3; TFS). PREVIEW version 2.3.5 (Protein Metrics Inc.) [41] was used to inspect the raw files to determine optimal search criteria, and the following search parameters were used: (a) enzyme specified as trypsin with maximum of two missed cleavages allowed; (b) acetylation of protein N-terminal including loss-of-methionine; (c) oxidation of methionine; (d) deamidation of asparagine/glutamine as dynamic post-translational modification; (e) carbamidomethylation of cysteine as static; Precursor mass-tolerance of 10 PPM while fragment mass-tolerance of 0.02 Dalton. PD's node, Spectrum file RC, was set up to query the raw files against the human proteome downloaded from UniProt (*Homo sapiens*, UP000005640, October 2018) with the static modification to recalibrate and detect features with the Minora node. Further, the internal contaminant database was also queried along with the human proteome using Sequest [42] search engine available in PD. For downstream analysis of these peptide-spectrum matches (PSM), both protein and peptide identifications/PSM false discovery rate (FDR) were set to 1%; thus, only the unique peptides with high confidence were used for final protein group identification. Peak abundances were extracted by integrating the area under the peak curve. Each protein group abundance was normalized by the total abundance of all identified peptides at  $\text{FDR} < 1\%$ . Summed up median values for all unique peptide ion abundances mapped to respective protein using label-free quantification scaled on all average with Precursor Ion Quantifier node [43] for PD were used.

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the data set identifier PXD024164.

The protocol for SILAC phosphoproteomic analysis was thoroughly described in the other study [26]. Briefly, protein sample equivalent to 4 mg was dissolved and fractionated, followed by phosphorylated peptide enrichment with MagReSyn-TiIMAC beads (Resyn Biosciences, Edenvale, Gauteng, South Africa) and Magnetic Rack (DynaMag-2; Life Technologies, Carlsbad, CA, USA). Analyses of peptides for total proteome and phosphorylated proteome were carried out on a Velos-Pro Orbitrap (TFS) mass spectrometer coupled with a Dionex UltiMate 3000 RS (TFS). The raw data files obtained from the mass spectrometric outputs for each experiment were merged into a single quantitated data set using MAXQUANT (version 1.5.2.8)



[45] and the ANDROMEDA search engine software [46]. The mass spectrometry phosphoproteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the data set identifier PXD025858.

### 2.3. Metabolomic analysis

The metabolomic experimental set-up was illustrated in Fig. S1A. Cells were washed with saline solution (0.9%), and the metabolites were extracted with MeOH : dH<sub>2</sub>O (80 : 20) containing an internal standard mix (Table S1). After adding MeOH : dH<sub>2</sub>O (80 : 20), samples were centrifuged, and the supernatant was taken and vacuum dried. The extracts were analysed on UPLC mass spectrometry (SYNAPT G2; Waters) according to published protocols [47]. Metabolites were identified and quantified in MASSLYNX software (version 4.2) from waters. For untargeted data analysis, XCMS [48] was used for automatic peak-picking (centWave) [49] and retention time alignment (OBI-Warp) [50]. All features that eluted in the first 66 s were omitted from further analysis. Feature intensities were normalized using quality control sample-based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC) [51] which was implemented using the R-package *NormalizeMets* [52]. For quality assurance, all features with over 25% relative standard deviation in the QC samples were omitted from further analysis. Generalized logarithmic transformation (glog) [53] and autoscaling were used to obtain mean-centred, normally distributed feature intensity values with equal variance. The expression of metabolites was normalized to cell numbers estimated by crystal violet assays. For the normalization of the metabolic measurements in the metabolomic experiment, cells were counted using a crystal violet assay. In short, cells were fixed with 100% cold MeOH and stained with 0.25% crystal violet (Merck; C.I. 42555). After washing, stained cells were dissolved into 100  $\mu$ L of 10% acetic acid and measured at 570 nm in the microplate reader (SpectraMax<sup>®</sup> M3; Molecular Devices LLC, San Jose, CA, USA).

### 2.4. siRNA transient knock-down and quantitative reverse transcription PCR

Cells were seeded either at 60 000 cells/well in 48-well plates or at 480 000 cells/well in 6-well plates. Before cell seeding, plates were coated with respective control siRNA (Silencer<sup>™</sup> Select Negative Control, 4390843) and target siRNA (Silencer<sup>™</sup> Select siUGDH: s409 and s410; siPDGFRB: s10240; siRELA: s11914 and s11915) as well as Lipofectamine<sup>™</sup> RNAiMAX

Transfection Reagent (TFS). Cells were transfected at 37 °C and 5% CO<sub>2</sub> for 48 h with the final siRNA concentration of 10 nM.

In the RT-qPCR experiments, cells were cultured in 48-well plates for 72 h, followed by total RNA extraction with TRI Reagent<sup>™</sup> Solution (Invitrogen<sup>™</sup>, TFS). RNA concentration was determined in NanoDrop One (TFS). 500–1000 ng of RNA was used for cDNA synthesis on the thermal cycler (Peltier Thermal Cycler, MJ research, PTC-225, Alameda, CA, USA) using High-Capacity cDNA Reverse Transcription Kit (TFS). Gene expression was measured with SYBR Green (Luna<sup>®</sup> Universal qPCR Master Mix; New England BioLabs, Ipswich, MA, USA) on Bio-Rad CFX384 Touch<sup>™</sup> Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Primers were either selected from Primer-Bank, designed on the Primer3Plus website, or based on the literature. Primer sequences for genes studied in this study were listed in Table S2.

### 2.5. Cell proliferation assay

Cells in quadruplicate were seeded at 10 000 cells/well in 96-well plates. *UGDH* knock-down followed the methods described above. For D492M, 24 h after seeding (48 h for D492HER2), cells were placed under the microscope (LEICA CTR 6500, bright field, 10 $\times$ ) at 37 °C with 5% CO<sub>2</sub> for real-time monitoring and multiple data acquisition. The microscope was controlled by software MICRO-MANAGER (version 1.4.22, Vale's laboratory, San Francisco, CA, USA). Three spots were chosen in each well, and photographs were taken every 6 h. Cell growth was monitored for 66 h for D492M while 42 h for D492HER2. Photographs were batch-processed with Macro in software IMAGEJ 1.52p (NIH, Bethesda, MD, USA), and cell numbers were normalized to the starting point.

### 2.6. Transwell invasion assay

The D492M and D492HER2 cells were cultured with siRNA transfection (Scramble and siUGDH) for 48 h in 6-well plates. *UGDH* knock-down followed the methods described above. Cells were then reseeded into filter units (Falcon<sup>®</sup> Permeable Support for 24-well Plate with 8.0- $\mu$ m Transparent PET Membrane, 353097, Corning, NY, USA) coated with Matrigel (Corning<sup>®</sup> Matrigel<sup>®</sup> Matrix, 356234) at a density of 30 000 cells/well. First, the filter inserts were coated with 100  $\mu$ L 1 : 10 diluted Matrigel for 20–30 min at 37 °C. Next, 300  $\mu$ L of cell suspension was added on top of the filter units. Then, 500  $\mu$ L of H14 medium with 10% FBS was added to the wells in the 24-well

plates below the filters. Finally, cells were incubated at 37 °C and 5% CO<sub>2</sub> for 48 h. Noninvasive cells on top of the filters were removed with cotton swabs, followed by fixation with paraformaldehyde (PFA, 3.7%, Sigma; 252549) and DAPI staining (1 : 5000; Sigma, D9542). Ten images per filter unit were taken by the EVOS<sup>®</sup> FL Auto Imaging System (10×; TFS), followed by the batch analysis of the images in Macro IMAGEJ 1.52p. For normalization of the different cell numbers in the filter units, cells were seeded into a 24-well plate along with the filter units and cultured and treated in the same way as cells in the filter units.

## 2.7. Statistical analysis and bioinformatics

All experiments performed in this study were in at least triplicates. The metabolomic analysis of the *UGDH* knock-down treatment was in six replicates. The proteomic data were processed in PERSEUS (version 1.6.14.0, data imputation based on normal distribution, width = 0.3, downshift = 1.8, permutation-based FDR < 0.05) [54] and R (version 4.0.0, the University of Auckland, New Zealand). Plots in this study were generated in R software. The statistical significance for all two-sample comparisons was based on the two-sided Student's *t*-test (Welsch, *P* < 0.05). Gene Ontology (GO) functional annotation was conducted in DAVID (DAVID Bioinformatics Resources 6.8) with default settings [55,56]. Reactome pathway analysis was performed with Reactome (Pathway browser version 3.7; Reactome database release: 75) with default settings [57]. Proteins with permutation-based FDR < 0.05 were used for the GO annotation and Reactome pathway analysis. Patient survival was plotted in KM plotter (kmplot.com) with basal breast cancer patients (split patients by autoselect best cut-off) [58]. The phosphoproteomic data were analysed in the INGENUITY PATHWAY ANALYSIS (IPA) (QIAGEN, Germantown, MD, USA, version from 2018) for pathway enrichment and PERSEUS for motif enrichment analysis.

All the R codes used for figure plotting in this study could be found on [https://github.com/QiongW56/UGDH\\_Publication\\_2021](https://github.com/QiongW56/UGDH_Publication_2021).

## 3. Results

### 3.1. The proteomic differences based on cell-of-origin outweigh proteomic changes that accompany EMT

Three breast EMT cell models consisting of epithelial and mesenchymal breast cell line pairs were used in this

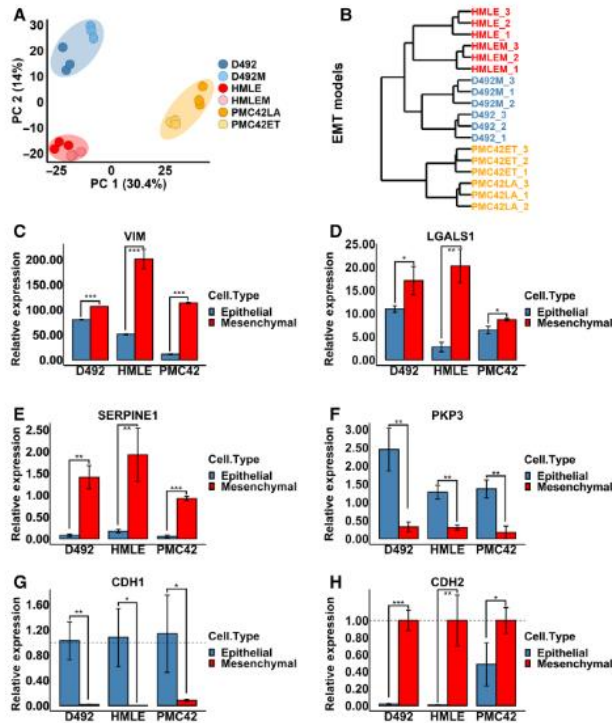
study (Fig. 1), namely, D492/D492M (D492 EMT model), HMLE/HMLEM (HMLE EMT model) and PMC42LA/PMC42ET (PMC42 EMT model). All three epithelial cell lines presented a typical cobblestone-shaped epithelial cell phenotype, while all the mesenchymal cells showed flattened mesenchymal morphology with undefined cell contour (Fig. S1B). The three EMT cell models presented different luminal/myoepithelial/basal phenotypes, with all three models possessing certain degrees of basal breast cell properties.

Irrespective of being epithelial or mesenchymal, cell lines of the same origin were grouped on the proteomic level (Fig. 2A). The PMC42 model shared the least similarities with the other EMT models (Fig. 2B). To confirm the epithelial and mesenchymal phenotypes on the molecular level, we quantified the EMT markers captured by the proteomic analysis from an EMT marker database [59]. VIM, LGALS1 and SERPINE1 were consistently upregulated in the mesenchymal cells, while PKP3 was downregulated (Fig. 2C–F). Not all EMT markers found in this study were, however, consistently altered among all three models, that is CD44, LMNB1, MSN, FLNA, TLN1, FSCN1, EGFR, S100A2 and NDRG1 (Fig. S2). Since E-cadherin (CDH1) and N-cadherin (CDH2), two typical EMT markers [60], were not covered in the proteomic analysis, we checked the expression of these by real-time PCR. CDH1 was significantly downregulated, while CDH2 was significantly upregulated in all EMT models (Fig. 2G,H).

### 3.2. Cell-cell and cell-extracellular matrix interactions are altered in EMT, and a diversity of pathways and molecular activities are changed in D492 as opposed to protein translation in HMLE and PMC42

Heterogeneity and plasticity are two intrinsic characteristics of EMT. To further define the epithelial and mesenchymal cells in all three EMT models, we compared their proteomes with respect to the number and profile of the significantly altered proteins along with their biological function and identified consistent EMT markers.

In total, 873 proteins were deemed valid proteins in identification and quantification (Table S3). In the D492 model, 188 out of the 873 valid proteins (21.5%) were significantly changed after EMT (permutation-based FDR < 0.05). In the HMLE model, 436 out of 873 proteins (49.9%) were significantly altered, while 200 proteins (22.9%) were significantly changed in the PMC42 model (Fig. 3A). Out of the significantly altered proteins, 55.9% (105/188) in the D492 model, 18.8% (82/436) in the HMLE model, while 63.5% (127/200) in the PMC42 model were upregulated after EMT



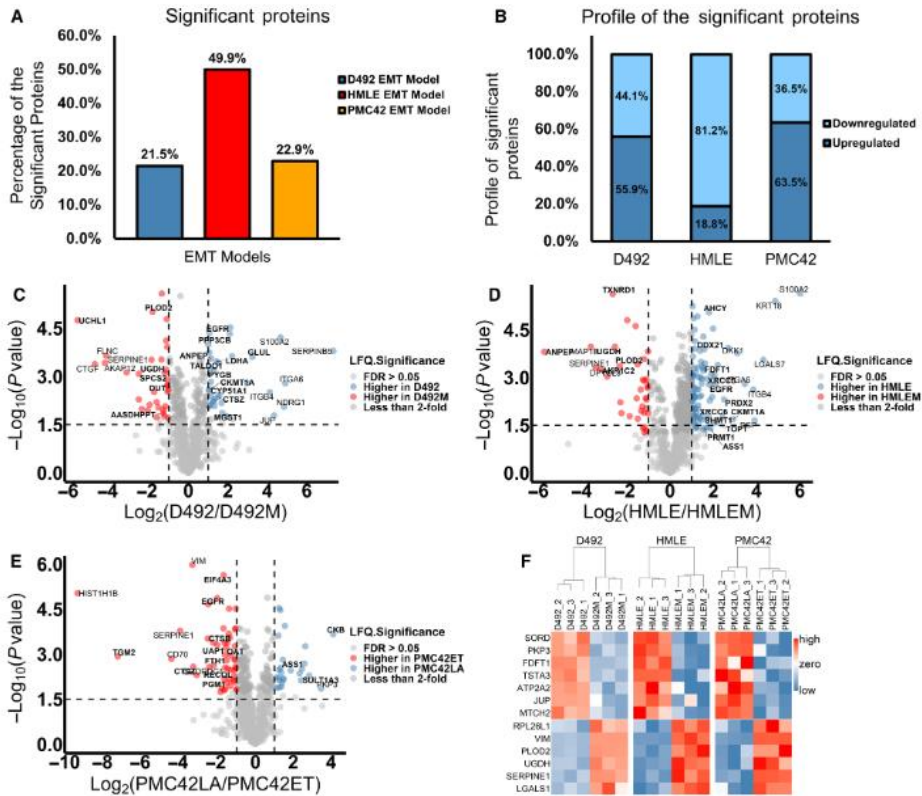
**Fig. 2.** EMT markers in the three breast EMT cell models. (A, B) Proteomic analysis of the three EMT models revealed cell lines with the same origin were more similar than their epithelial or mesenchymal states. The D492 EMT model was more similar to the HMLE model than the PMC42 model. Proteins with valid identification and quantification were included in the analysis (Table S3). Protein levels of the known EMT markers VIM (C), LGALS1 (D), SERPINE1 (E) and PKP3 (F) were consistently altered in all EMT models. RNA expression of CDH1 was downregulated (G), while RNA expression of CDH2 was upregulated after EMT (H). Student's *t*-test, \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ;  $n = 3$ . The error bars indicate standard deviation. VIM, vimentin; LGALS1, galectin-1; SERPINE1, plasminogen activator inhibitor 1; PKP3, plakophilin-3; CDH1, E-cadherin; CDH2, N-cadherin.

(Fig. 3B). To ensure reproducibility of the proteomic data used in this study (Table S4), we compared the current proteomic data set with the previously generated data for the D492 EMT model [26]. The correlation coefficient of these two data sets was 0.936 (Fig. S3).

We next filtered the identified proteins based on their log<sub>2</sub> fold changes and  $-\log_{10}$  *P*-values (Fig. 3C–E) and summarized the consistently altered proteins in all three EMT models (Fig. 3F and Table 1) to identify common changes in EMT. A literature search for each EMT target revealed that all had been associated with EMT previously, albeit to a different extent (Table 1). To evaluate these consistently altered EMT markers in the context of cancer progression, we

confirmed the expression of these markers in the tumorigenic breast mesenchymal cell line D492HER2. All the targets detected in D492HER2 showed the same trends in changes (Table 1).

To define functional changes in EMT, we annotated the GO terms for the significantly changed proteins (Table S3) and observed that the Biological Process (BP) 'cell-cell adhesion' was altered in all three EMT models (Fig. 4A–C). The D492 model had the least similarity compared with the other two models, with only one common BP term (i.e. cell-cell adhesion) out of the top 10 enriched BP terms (Fig. 4A). In contrast, the PMC42 model shared its top seven terms with HMLE (Fig. 4B, C). The same trend was observed using enriched



**Fig. 3.** Proteomic analysis of the three breast EMT cell models. (A) Percentages of significantly altered proteins in the EMT models (permutation-based FDR < 0.05). (B) Up- and downregulation profiles for all the significantly changed proteins in the EMT models. The log<sub>2</sub> (epithelial/mesenchymal ratio) along with the  $-\log_{10}(P\text{-value})$  for each protein was plotted for the D492 model (C), HMLE model (D) and PMC42 model (E). Proteins with FDR (permutation-based) less than 0.05 and fold change more than 2 are coloured. The horizontal and vertical dashed lines indicate a  $P\text{-value}$  of 0.03 [ $-\log_{10}(P\text{-value}) = 1.5$ ] and a 2-fold change, respectively. The annotated proteins had a log<sub>2</sub> (fold change) of more than 3. Proteins involved in metabolism with a log<sub>2</sub>(fold change) of more than 1 for D492 model, 1.5 for HMLE and PMC42 models were bold label marked.  $n = 3$ . (F) A list of proteins significantly changed in the same direction (Student's  $t\text{-test}$ ,  $P < 0.05$ ) in all three EMT models. SERPINE1, RPL26L1, PLOD2, UGDH, LGALS1 and VIM were upregulated, while JUP, PKP3, MTCH2, ATP2A2, FDF11, SORD and TSTA3 were downregulated after EMT.

Reactome pathway analysis (Fig. 4D–F). The altered Reactome pathways in the D492 model were related to response to cell stress, IGF signalling and interleukin-12 signalling (Fig. 4D). In both the HMLE and PMC42 models, changes were, however, mainly to pathways involved in the protein translational process (Fig. 4E,F). Comparison of changes to cellular components (Fig. S4A–C) and molecular function (Fig. S4D–F) was similarly indicative of more similarities in changes to

protein function following EMT in the HMLE and PMC42 models as compared to the D492 model.

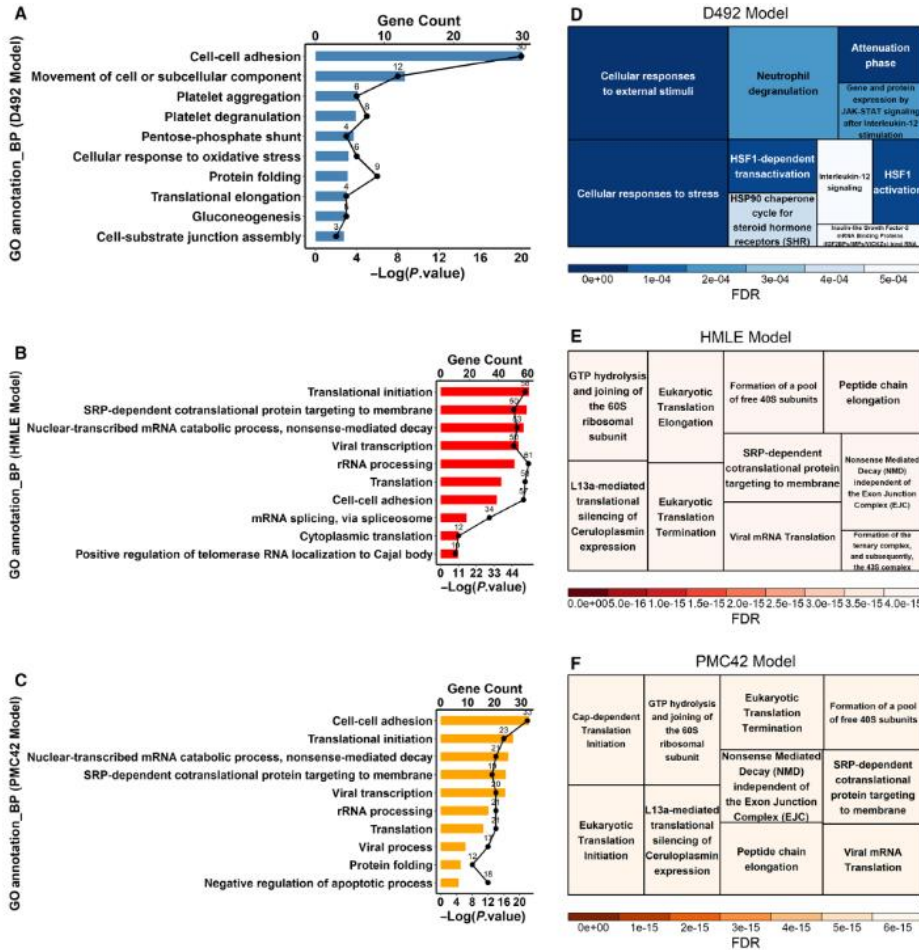
### 3.3. UGDH is negatively correlated with patient survival and affects cell proliferation, cell invasion and SNAI1 expression

Next, we focused on the metabolic changes during EMT. Out of the thirteen identified targets listed in



**Table 1.** The EMT targets were significantly different (Student's *t*-test,  $P < 0.05$ ) in all EMT models. The literature related to each target in terms of EMT was also listed. Changes in these targets in another breast mesenchymal cell line with tumorigenicity were consistent with the findings in this study.

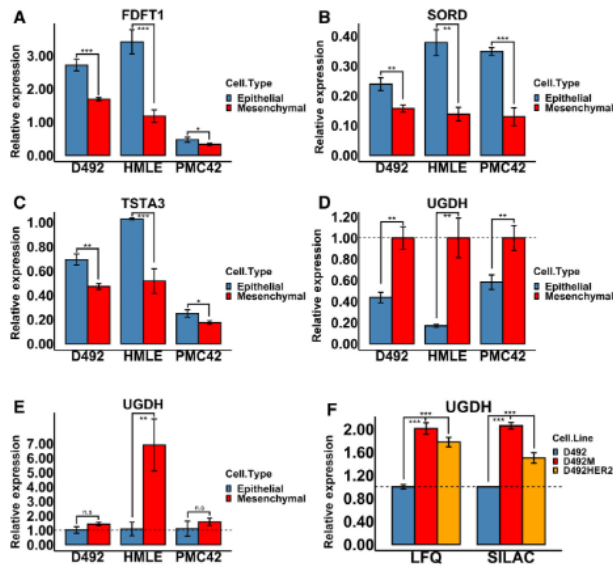
UniProt ID	Gene name	Protein name	Gene name	Log2 (D492/ D492M)	Log2 (HMLE/ HMLEM)	Log2 (PMC42LA/ PMC42ET)	-Log (P-value D492s)	-Log (P-value HMLEs)	-Log (P-value PMC42s)	EMT-related literature	Gene expression in mesenchymal cells with tumorigenicity (Log2 ratio)
P05121-1	Plasminogen activator inhibitor 1		<i>SERPINE1</i>	-4.236	-3.442	-3.993	3.416	3.313	3.780	[70,72,97]	-4.739
Q9JUN3	60S ribosomal protein L25-like 1		<i>RPL26L1</i>	-2.367	-1.123	-1.346	1.981	2.349	1.925	[98]	Not detected
O00469-1	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2		<i>PLOD2</i>	-1.817	-1.504	-1.121	5.005	3.411	2.072	[70,71]	-1.687
O60701	UDP-glucose 6-dehydrogenase		<i>UGDH</i>	-1.193	-2.550	-0.779	3.101	3.982	2.285	[27,28,32,34,75]	-0.828
P09382	Galectin-1		<i>LGALS1</i>	-0.623	-2.912	-0.444	1.769	3.035	1.799	[67,69,70]	-0.422
P08670	Vimentin		<i>VIM</i>	-0.407	-1.979	-3.343	5.493	4.802	5.970	[4]	-0.957
Q13630	GDP-L-fucose synthase		<i>TSTA3</i>	0.553	1.009	0.508	2.871	2.517	1.789	[99,100]	0.562
Q00796	Soritol dehydrogenase		<i>SORD</i>	0.599	1.459	1.452	2.457	2.957	2.806	[101]	0.111
P37268	Squalene synthase		<i>FDFT1</i>	0.683	1.536	0.491	3.510	3.215	1.462	[102]	0.722
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2		<i>ATP2A2</i>	0.919	0.523	0.641	2.388	1.551	2.316	[103]	0.299
Q9Y6C9	Mitochondrial carrier homolog 2		<i>MTC2</i>	1.239	0.725	0.899	2.219	1.624	3.015	[99]	0.198
Q9Y446	Plakophilin-3		<i>PKP3</i>	2.968	2.057	3.478	2.766	3.077	1.850	[66,68]	3.458
P14923	Junction plakoglobin		<i>JUP</i>	4.343	1.541	1.214	1.783	1.358	3.611	[104,105]	4.344



**Fig. 4.** Functional annotation of the GO terms (BP) and Reactome pathway analysis for the three EMT models. (A–C) Functional annotation of the GO terms (BP) was conducted on the DAVID platform (DAVID Bioinformatics Resources 6.8) for the D492 model (A), HMLE model (B), and PMC42 model (C). The GO terms were listed according to the  $-\log_{10} P$ -value in descending order. The numbers of genes in each GO term were also plotted as dots/line plots. (D–F) Reactome pathway analysis (Pathway browser version 3.7; Reactome database release: 75) for the D492 model (D), HMLE model (E) and PMC42 model (F). Data used for both the GO annotation and the pathway analysis (Table S3) were proteins significantly different in each EMT model (permutation-based FDR < 0.05). Default settings in the DAVID and Reactome platforms were used. BP, biological process.

Table 1, four proteins were involved in metabolism: FDF1, SORD and TSTA3 were downregulated while UGDH was upregulated (Fig. 5A–D). We further tested the RNA expression of UGDH, which showed

the most changes to protein expression in all EMT models and was associated with cancer aggressiveness. Though there was no significance in the D492 and PMC42 models, the upregulating trends in all EMT



**Fig. 5.** Four metabolic enzymes changed consistently in all EMT models. (A–D) The proteomic analyses revealed that the metabolic enzymes FDF11, SORD, TSTA3 and UGDH changed consistently in all EMT models. (E) The RNA level of UGDH in all EMT models was consistently higher in the mesenchymal cell lines. (F) The UGDH protein level in the epithelial and mesenchymal cells was confirmed in another data set [26] and further confirmed in another tumorigenic breast mesenchymal cell line D492HER2. Student's *t*-test, \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ;  $n = 3$ . The error bars indicate standard deviation. FDF11, squalene synthase; SORD, sorbitol dehydrogenase; TSTA3, GDP-4-fucose synthase; UGDH, UDP-glucose 6-dehydrogenase.

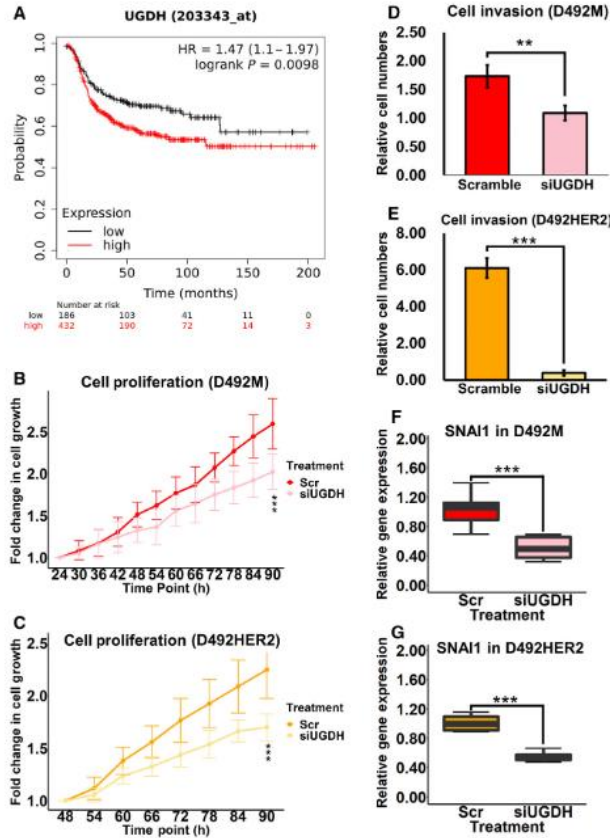
models were seen (Fig. 5E). To relate these findings to breast cancer, we tested the protein level of UGDH in the tumorigenic breast mesenchymal cell line D492HER2. UGDH was upregulated in D492HER2 as observed in nontumorigenic mesenchymal cell line D492M (Fig. 5F).

Recent studies have reported that UGDH affects patient survival [34], cell proliferation [32,37], cell invasion [27], cell migration [34,37] and SNAIL expression [28]. We set out to confirm these effects of UGDH in our EMT cell lines. High UGDH level was associated with worse patient survival in basal breast cancer patients based on KM plotter (Fig. 6A). Based on this, we analysed effects of UGDH on cell morphology, proliferation, invasion and SNAIL expression in two types of breast mesenchymal cells: nontumorigenic D492M and tumorigenic D492HER2 via siRNA-mediated knock-down of *UGDH* (Fig. 5S). Knock-down of *UGDH* did not yield observable morphological changes but slowed down cell growth (Fig. 6B,C) and invasion (Fig. 6D,E and Fig. S6A–C) in both cell lines.

SNAIL RNA expression was downregulated after *UGDH* knock-down, which was also consistent with the literature (Fig. 6F,G and Fig. S6D,E).

### 3.4. GPC is downregulated while NAA is upregulated following *UGDH* knock-down in the mesenchymal cells

UGDH catalyses the conversion of UDP-Glc to UDP-GlcA that are constituents of glycosaminoglycans and N- and O-linked glycans [61]. To confirm the metabolic impacts of UGDH in mesenchymal cells, we knocked down *UGDH* with siRNAs and performed metabolomic analysis in all three mesenchymal cell lines. Samples from the same cell line clustered together at the metabolic level despite *UGDH* knock-down (Fig. 7A,B). As with the proteome, the metabolome of D492M was closer to that of HMLEM than the metabolome of PMC42ET. Knock-down of *UGDH* did not confer a distinct metabolic phenotype compared with the scramble control in any of the

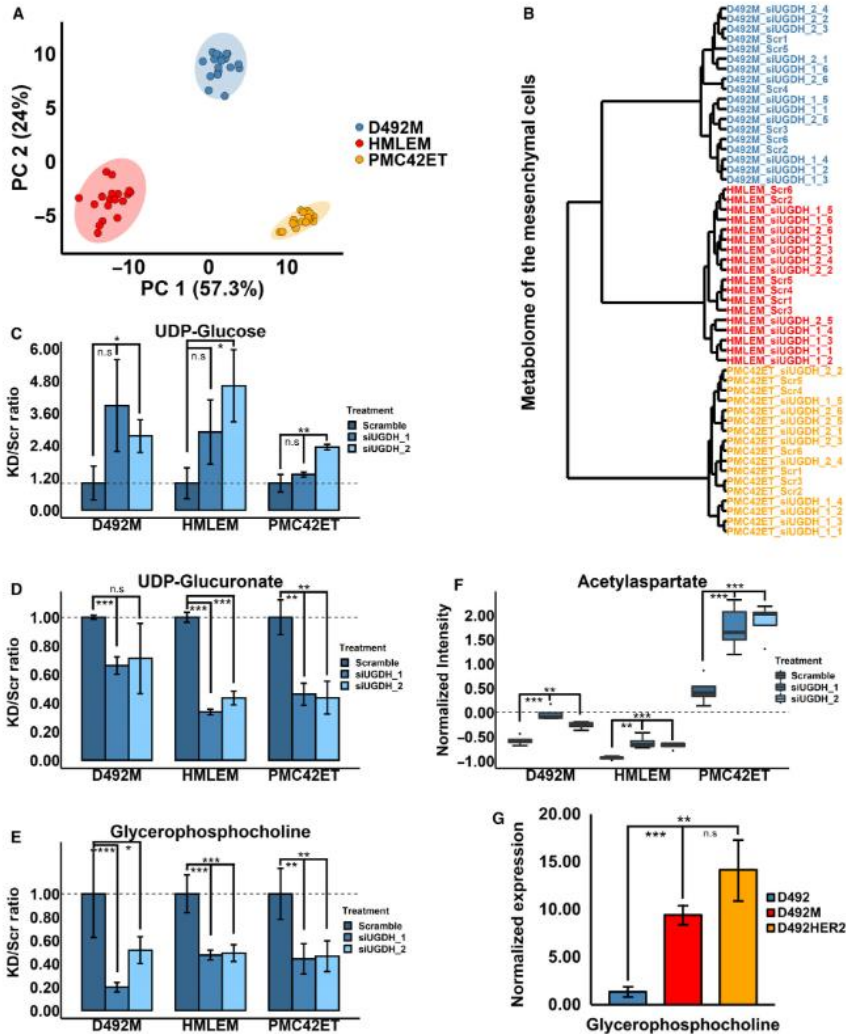


**Fig. 6.** Functional analysis of UGDH in EMT. (A) The Kaplan–Meier plot of UGDH in basal breast cancer patients was downloaded from kmpplot.com. (B, C) Cell proliferation slowed down with the siRNA knock-down of *UGDH* in both nontumorigenic D492M (B) and tumorigenic D492HER2 (C). *n* equals 4, and three spots were chosen for each replicate during the imaging process. (D, E) Cell invasion decreased with *UGDH* knock-down in both nontumorigenic D492M (D) and tumorigenic D492HER2 (E). *n* equals 3, and 10 spots were chosen for each replicate during the cell counting process. (F, G) One of the main EMT transcription factors SNAI1 was downregulated after the siRNA knock-down of *UGDH* in both nontumorigenic D492M (*n* = 7) (F) and tumorigenic D492HER2 (*n* = 4) (G). Student’s *t*-test, \*\**P* < 0.01; \*\*\**P* < 0.001. The error bars indicate standard deviation. SNAI1, Snail family transcriptional repressor 1; UGDH, UDP-glucose 6-dehydrogenase.

mesenchymal cell lines (Fig. 7B). An increasing trend of UDP-Glc was observed in all the mesenchymal cell lines with all the siUGDH treatments, although non-significant for one of the siRNAs (Fig. 7C). UDP-GlcA decreased in all the mesenchymal cell lines in all the siUGDH treatments, although nonsignificantly with one siRNA in D492M (Fig. 7D).

To better evaluate the systemic changes of UGDH on metabolism, we carried out an untargeted metabolomics analysis. Knocking down *UGDH* significantly decreased the intracellular glycerophosphocholine (GPC) level and increased acetylaspartate (NAA) in all the mesenchymal cell lines (Fig. 7E,F), which was confirmed in the aggressive D492HER2 and MDA-





**Fig. 7.** Metabolic changes after siRNA knock-down of *UGDH* in the mesenchymal cell lines. (A, B) Metabolomic clustering of the mesenchymal cell lines with different treatments. Valid metabolite identification and quantification from the negative, positive and basic modes were integrated into the analysis. Samples clustered together based on the differences of their metabolome among different EMT models (A), and the D492 mesenchymal cells were closer to the HMLEM mesenchymal cells than PMC42 at the metabolic level (B). (C) The *UGDH* substrate UDP-glucose was increased following siRNA knock-down of *UGDH* in all cell lines with two siRNAs ( $n = 3$ ). (D) The *UGDH* product UDP-glucuronate was decreased with siRNA knock-down of *UGDH* in all the cell lines with two siRNAs ( $n = 3$ ). (E, F) siRNA knock-down of *UGDH* significantly decreased GPC ( $n = 5$ ) and increased acetylaspartate ( $n = 4$ ) in all the cell lines with two siRNA treatments. (G) GPC level was higher in the nontumorigenic D492M than the epithelial D492, and it was the highest in the tumorigenic mesenchymal D492HER2 ( $n = 3$ ). Student's *t*-test, \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . The error bars indicate standard deviation. *UGDH*, UDP-glucose 6-dehydrogenase.

MB-231 cell lines (Fig. S7A–D). To investigate whether GPC and NAA were associated with the UGDH level and differently expressed regardless of tumorigenicity, we tested the GPC and NAA levels in the epithelial D492, nontumorigenic mesenchymal D492M and tumorigenic mesenchymal D492HER2 cells. GPC was higher in both D492M and D492HER2 compared with D492 (Fig. 7G). We further looked into the connection between GPC and the mesenchymal state based on published data sets in the literature [62,63] but did not observe any significant correlation (Fig. S7E and Table S5). siRNA-mediated knock-down of *UGDH* did not yield significant and consistent changes to choline and phosphocholine (Fig. S7F,G).

Glycerophosphocholine is part of the choline synthetic pathway from phosphatidylcholine (PtdCho), and NAA is closely associated with acetyl-CoA and central carbon metabolism. To query how changes to UDP-GlcA might relate to GPC and NAA processing via changes in metabolic flux, we performed *in silico* knock-down of *UGDH* using tailored genome-scale metabolic models of D492 [23,25]. Changes to metabolic flux were observed within keratan metabolism, hyaluronan processing, pentose phosphate pathway and the central carbon metabolic pathways (Table S6). Negligible changes were, however, observed to GPC production and consumption.

### 3.5. PDGFRB signalling regulates UGDH potentially via NFκB-p65

We next investigated the upstream regulation of UGDH by analysing the secretome of the D492 model [64]. IGF, TGF-β and PDGFD signalling regulators were highly presented in the culture medium of D492M cells (Fig. 8A). PDGFRB was highly expressed in the nontumorigenic D492M (Fig. 8B) and the tumorigenic D492HER2 mesenchymal cell lines (Fig. S8A) [65], and PDGFD was secreted by D492M (Fig. 8C). We thus focused on the role of PDGF

signalling in UGDH regulation. In addition, the motif enrichment analysis of the phosphorylation sites within the phosphoproteomic data (Table S7) revealed potentially altered kinases in the D492 EMT model, including the downstream target of the PDGF signalling PKC kinase (Fig. 8D). siRNA-mediated knock-down of *PDGFRB* decreased both the PDGFR signalling downstream regulator *RELA* (NFκB-p65) and UGDH in D492M (Fig. 8E–G) and D492HER2 (Fig. S8B–D). We further investigated the impact of *RELA* on UGDH and found that siRNA-mediated knock-down of *RELA* decreased the UGDH RNA level in D492M (Fig. 8H,I and Fig. S8E,F). We observed the same effect of *RELA* knock-down on UGDH in D492HER2 with only one siRNA (Fig. S8G–J).

## 4. Discussion

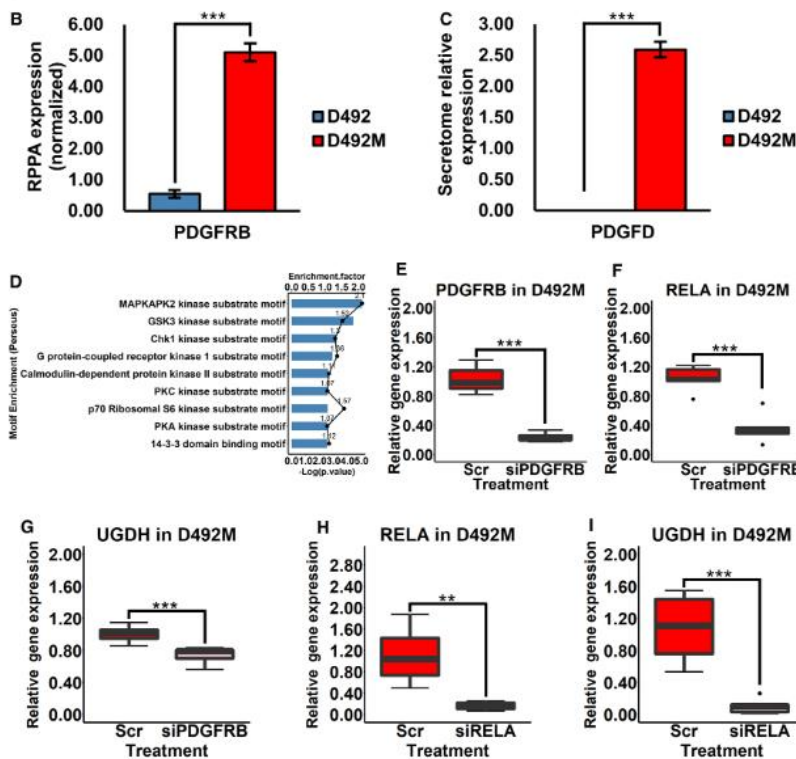
Herein we set out to determine common metabolic changes in cell models used to study EMT in breast epithelium. We chose the D492, HMLE and PMC42 EMT cell models on account of the spontaneous EMT induction approaches and the nontumorigenic properties of these cell lines to ensure the intrinsic characteristics and plasticity of EMT (Fig. 1). First, we validated and compared the EMT cell models on the proteomic level (Fig. 2A,B). Cell lines clustered based on their origin instead of their epithelial or mesenchymal characteristics, indicating that the spontaneous epithelial–mesenchymal switches during EMT/MET are subtle compared with the imprinted intrinsic genetic differences among these cell models.

VIM, LGALS1, SERPINE1, PKP3 and the CDH1-CDH2 switch were consistently altered in all the EMT models (Fig. 2C–H) and have all been related to EMT in different cancer types [4,60,66–72]. Vimentin, a type III intermediate filament and well-known EMT marker, shapes the cell structure and modifies cell movements and cell adhesion [73]. SERPINE1, a key player in endothelial homeostasis, is highly upregulated in EMT. However, the function of SERPINE 1 in EMT

**Fig. 8.** PDGFRB regulates UGDH via *RELA* (NFκB-p65) in D492M. (A) Top differently secreted growth factors from the secretome of the D492 EMT model were reported (permutation-based FDR < 0.05). (B) PDGFRB was highly expressed in mesenchymal cells than epithelial cells in the D492 model on the protein level based on the RPPA analysis ( $n = 3$ ) [65]. (C) PDGFD protein was highly secreted in mesenchymal cells than epithelial cells in the D492 model ( $n = 3$ ). (D) Motif enrichment of the phospho-proteome in the D492 EMT model suggested that PKC kinase activity, among others, was highly enriched in EMT. Enrichment factors  $\geq 2$ ; motif enrichment terms were ranked based on the  $-\log_{10}(P\text{-value})$ . (E) The knock-down efficiency of *PDGFRB* with siRNA in the D492M cell line was around 80% ( $n = 7$ ). (F) *RELA* (NFκB-p65) was downregulated after the siRNA knock-down of *PDGFRB* in D492M ( $n = 5$ ). (G) UGDH was downregulated after the siRNA knock-down of *PDGFRB* in D492M ( $n = 7$ ). (H) The knock-down efficiency of *RELA* with the first siRNA in D492M was about 80% ( $n = 6$ ). (I) UGDH was downregulated after the siRNA knock-down of *RELA* in D492M with the first siRNA ( $n = 6$ ). Student's *t*-test, \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . The error bars indicate standard deviation. PDGFRB, platelet-derived growth factor receptor beta; *RELA* (NFκB-p65), nuclear factor NF-κappa-B p65 subunit; UGDH, UDP-glucose 6-dehydrogenase.

**A**

Protein ID	Protein Name	Gene Name	-Log(Pvalue)	FDR	-Log2(D492/D492M)
Q8WX77	Insulin-like growth factor-binding protein-like 1	IGFBPL1	2.633	0.004	-6.055
P22692	Insulin-like growth factor-binding protein 4	IGFBP4	6.401	0.000	-5.008
P61812	Transforming growth factor beta-2	TGFB2	4.302	0.000	-4.524
P01137	Transforming growth factor beta-1	TGFB1	1.773	0.014	-3.392
Q16270	Insulin-like growth factor-binding protein 7	IGFBP7	5.191	0.000	-3.341
Q9GZP0	Platelet-derived growth factor D	PDGFD	4.127	0.000	-3.273
P18065	Insulin-like growth factor-binding protein 2	IGFBP2	2.878	0.003	-1.729
Q15582	Transforming growth factor-beta-induced protein ig-h3	TGFB1	2.606	0.004	-0.967
M0QXF7	Myeloid-derived growth factor (Fragment)	MYDGF	3.098	0.002	-0.848
G3V3X5	Latent-transforming growth factor beta-binding protein 2	LTBP2	2.411	0.005	-0.750
J3KT38	Growth factor receptor-bound protein 2 (Fragment)	GRB2	1.933	0.013	-0.295
H7BYW6	Platelet-derived growth factor subunit A (Fragment)	PDGFA	1.646	0.019	0.505
P08581	Hepatocyte growth factor receptor	MET	1.478	0.024	0.610
P51858	Hepatoma-derived growth factor	HDGF	1.829	0.014	0.647
AAOAC4DH07	Latent-transforming growth factor beta-binding protein 4	LTBP4	2.645	0.004	0.827
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	2.810	0.003	1.728
Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2	EPS8L2	3.919	0.000	1.835
P00533	Epidermal growth factor receptor	EGFR	2.168	0.010	2.485
C9JD84	Latent-transforming growth factor beta-binding protein 1	LTBP1	4.083	0.000	2.501
Q14512	Fibroblast growth factor-binding protein 1	FGFBP1	6.742	0.000	8.075



is poorly understood. The possible role of SERPINE1 in EMT is to affect the function of urokinase-type plasminogen activator receptor (uPAR) to regulate extracellular matrix degradation [72]. LGALS1 is a carbohydrate-binding protein. One study shows that upregulation of LGALS1 decreases CDH1 and increases SNAI1 [67]. PKP3 is an epithelial marker and is under the control of the EMT transcriptional regulator ZEB1 [66,68]. All these EMT markers were consistently altered in the three EMT models (Fig. 2C–F). However, inconsistencies in EMT markers were also observed indicative of their different roles in EMT with respect to cell type. The PMC42 model was different from the other EMT cell models (Fig. S2), potentially reflecting the cell heterogeneity and partially expressed mesenchymal marker CDH2 in the epithelial cells (Fig. 2H) [74]. The consistently altered EMT markers were also confirmed in the tumorigenic breast mesenchymal cell line D492HER2 (Fig. 3F and Table 1), indicating that these markers are not only crucial for EMT but also potentially involved in tumorigenicity and malignancy, even though they are not critical for tumour initiation. Moreover, many of the consistently altered proteins identified in this study remain unexplored in the context of EMT (Table 1).

Our findings confirmed that changes to cellular morphology, cell–cell communication and cell–extracellular matrix interaction are among the main characteristics of EMT (Fig. 4 and Fig. S4). Even though the proteomes of the D492 cell lines were closer to HMLE (Fig. 2B), they shared the least similarity in the altered pathways post-EMT. The changed translational activities in HMLE and PMC42 and the altered responses to stress and signalling regulation in D492 suggest that in HMLE and PMC42, the epithelial or mesenchymal switch may largely be mediated by altered expression of proteins involved, whereas in D492, post-translational control of existing proteins may play a more important role. This may also reflect the more stem-like properties of the D492 epithelial cells that confer cell flexibility. Our findings indicate that distinct and dominant cell properties (e.g. stem cell properties) outweigh similar genetic backgrounds for EMT induction, while cells with disparate genetic backgrounds can rely on similar machineries to induce EMT.

Recently, a growing number of studies have focused on UGDH in cancer, and the roles of UGDH in tumour growth, metastasis and patient survival have been well documented [27,32,34–37]. Additionally, UGDH has been connected to EMT [27,28,32,34,75]. Arnold *et al.* [27] reported that UGDH was highly expressed in mesenchymal cells and mesenchymal-like

breast cancers and connected UDP-GlcA (the enzymatic product of UGDH) to extracellular matrix remodelling and mesenchymal-like properties. Furthermore, UGDH regulates SNAI1, a well-known EMT transcription factor, via UDP-Glc (the enzymatic substrate of UGDH) [28]. We confirmed the upregulation of UGDH in both nontumorigenic and tumorigenic mesenchymal cell lines, suggesting UGDH is associated with the mesenchymal feature in tumorigenic cell lines (Fig. 5F). Interestingly, even though the high expression of UGDH was associated with worse survival in basal breast cancer patients (Fig. 6A) and decreased UGDH jeopardized cell proliferation (Fig. 6B,C) and invasion (Fig. 6D,E and Fig. S6A–C), all the mesenchymal cells in this study possess upregulated UGDH and are nontumorigenic. Thus, elevated UGDH expression is likely not a trigger for tumour initiation, but tumorigenic cells may rely on UGDH to facilitate tumorigenicity and malignancy. UGDH may induce resistance to chemotherapy via drug elimination. This was supported by a recent study demonstrating that high levels of UGDH are correlated with worse prognosis in triple-negative breast cancer patients receiving chemotherapy, likely by promoting UDP-GlcA-mediated detoxification and elimination of epirubicin [76]. The effect of UGDH on SNAI1 supports that UGDH has a regulatory role in EMT and that its function may exceed its catalytic role, perhaps via nonconventional signalling regulatory effects such as glycosylation (Fig. 6F,G and Fig. S6D,E).

The D492 EMT model metabolome was more similar to HMLE than PMC42 (Fig. 7A,B), consistent with the proteomic analysis (Fig. 2A,B). In agreement with the literature, knock-down of *UGDH* increased UDP-Glc and decreased UDP-GlcA (Fig. 7C,D), both of which are important metabolites with wide impact on cells [27,28]. The most prominently altered metabolite was, however, GPC (Fig. 7E). Increased GPC in tumours indicates changes to choline metabolism, which has emerged as a hallmark of cancer progression [77]. GPC is negatively correlated with patient survival [78] and is high in basal-like breast cancer xenograft and oestrogen receptor-negative breast cancer patients [79,80]. Reduced GPC levels after chemotherapies are associated with better survival in breast cancer patients [78]. D492 and D492M are basal-like breast cell lines, while D492HER2, deemed as HER2-positive breast cell line, is more closely associated with the aggressive claudin low than other breast cancer types [26,81]. Claudin low is not a distinct intrinsic breast tumour subtype but may permeate various breast cancer types including HER2-positive [82]. The higher levels of GPC along with UGDH in



basal-like mesenchymal D492M and claudin-low D492HER2 are in congruence with the clinical observations. GPC may be involved in EMT, but the connection between GPC and EMT in cancer is unclear [83,84]. Li and colleagues detected GPC in 928 cell lines and performed different types of metabolite-gene association analyses. They reported various genes associated with GPC where the EMT master regulator TWIST1 was one of the top hits [62]. The insignificant correlation between GPC and mesenchymal cells (Fig. S7E) suggests that the increased GPC levels observed in the D492 mesenchymal cells are results of one or several regulators independent of mesenchymal traits. Our results support that GPC in part is regulated by the mesenchymal metabolic enzyme UGDH, but the molecular mechanisms underlying this warrant further investigation. *In silico* knock-down of *UGDH* in the genome-scale metabolic models revealed several metabolic changes (Table S6) primarily on account of rerouting of glucose flux away from UDP-GlcA formation and into glycolysis and associated pathways (e.g. PPP and TCA), which may be partially responsible for the increased NAA (Fig. 7F). It, however, imparted no changes in GPC, implying that the changes to metabolic fluxes encircling GPC due to a mass-action effect of UGDH are likely secondary to changes that arise through altered glycosylation. Cell osmotic pressure balance is vital for normal cell functions and cell survival. GPC is a well-known intracellular osmotic regulator, and proteoglycans serve as extracellular osmolytes [85]. The decreased intracellular GPC may thus counterbalance the decreased extracellular osmotic pressure induced by the reduced proteoglycans caused by the knock-down of *UGDH*.

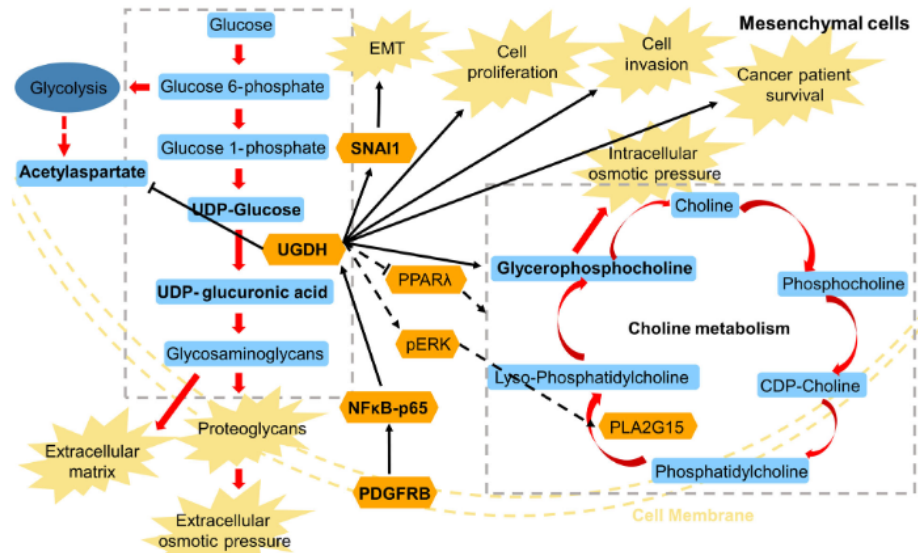
Recently, studies have shown that UGDH regulates signalling factors and lipid metabolic genes, such as SNAIL-, SIP-1-, ERK/MAPK-, SIX1- and PPAR $\gamma$ -targeted genes [27,28,32,34]. PPAR $\gamma$  is a nuclear transcription factor regulating genes linked to lipid metabolism [86,87] that interacts with choline/PtdCho metabolism [88]. UGDH has been proposed to inhibit PPAR signalling and affect lipid metabolism [27]. Consistent with this, we observed a negative association between UGDH expression and PPAR $\gamma$  signalling (Fig. S7H), suggesting the *UGDH* knock-down decreases intracellular GPC level via PPAR $\gamma$ . Moreover, phospholipase A2 group XV (PLA2G15), which belongs to Cytosolic phospholipase A2 (cPLA2), is an enzyme catalysing the hydrolysis of phospholipids, potentially involved in the formation of GPC from PtdCho, and is under the control of ERK signalling [89,90]. Knock-down of *UGDH* has been reported to downregulate the phosphorylation of ERK (pERK) in

highly invasive ovarian cancer cells [32]. We observed that both GPC and PLA2G15 were higher in the mesenchymal cell lines D492M and D492HER2 (Fig. 7G and Fig. S7I), implying UGDH may regulate GPC via pERK-PLA2G15. Taken together, UGDH may indirectly affect GPC via signalling regulations and/or lipid metabolism to retain the osmotic balance across the cell membrane, although further investigations are needed. Furthermore, the absence of UGDH in the list of genes associated with GPC reported in the literature indicates UGDH may not be a dominant GPC regulator [62].

Slit2, SP1, TGF- $\beta$ , hypoxia, p38<sup>MAPK</sup>, LMP2A and PI3K/Akt affect and/or regulate UGDH expression, which highlights that UGDH is under complex regulation network control [29–31,33]. These regulators are potentially mediated by PDGF signalling that, along with the downstream transcription factor NF $\kappa$ B, is dysregulated in cancer progression and EMT [7,91,92]. Tam *et al.* [7] reported a switch from EGFR to PDGFR signalling in cancer stem cell formation and EMT. The higher expression of PDGFRB and secretion of PDGFD in D492M compared with D492 suggest PDGFRB signalling is upregulated in mesenchymal cells (Fig. 8A–C and Fig. S8A) supported by the increased phospholipase C, PI3K/Akt and PKC $\alpha$  signalling (Fig. S7H and Fig. 8D) since these are well-known downstream targets of PDGFR [7,93]. This is consistent with PDGFD-PDGFRB signal regulation of EMT [94,95]. NF $\kappa$ B-p65 is a downstream regulator of PDGFR signalling [92,96]. Downregulating either *PDGFRB* or *NF $\kappa$ B-p65* decreased UGDH expression on the RNA level in both D492M and D492HER2. However, the impacts of PDGFRB and NF $\kappa$ B-p65 on UGDH were dampened in D492HER2 compared with D492M (Fig. 8E–I and Fig. S8B–J). We have previously noticed that D492HER2 is a less complete mesenchymal cell line than D492M, suggesting that the regulations of PDGFRB and NF $\kappa$ B-p65 on UGDH are more dominant in complete mesenchymal cells [26]. It thus appears that UGDH is part of an interactive signalling and metabolic network in which PDGFRB differently regulates UGDH via NF $\kappa$ B-p65 depending on specific cell types.

## 5. Conclusions

In conclusion, we used three breast EMT cell models to study proteomic changes in EMT, focusing on metabolic reprogramming. We further studied the downstream functions of the metabolic enzyme UGDH in cancer progression and metabolism, and finally, we explored the upstream signalling regulating UGDH (Fig. 9). Several proteins were found to be



**Fig. 9.** Summary of the study. This figure illustrates the main metabolic pathways and findings involved in this study. UGDH catalyses UDP-glucose into UDP-glucuronic acid, an indispensable unit for glycosaminoglycans, proteoglycans and extracellular matrix. In this study, UGDH was found to be highly expressed in the mesenchymal cells and affect cancer patient survival, mesenchymal cell proliferation and invasion and the EMT transcription factor SNAI1, and it was under the control of the PDGFRB-NFκB pathway. In addition, the knock-down of *UGDH* with siRNAs significantly decreased the intracellular GPC levels and increased the acetylaspertate (NAA) levels in all the mesenchymal cell lines. NAA is closely linked to the central carbon metabolism and potentially affected by the mass-action effects of *UGDH* knock-down. GPC is an intracellular osmolyte and part of the choline metabolism. Knock-down of *UGDH* hindered the formation of proteoglycans and further decreased the extracellular osmotic pressure, which could be counteracted by the reduced intracellular osmotic pressure induced by GPC. Based on the literature, we hypothesized that to ease the osmotic stress, knock-down of *UGDH* affected PPAR $\gamma$ -lipid metabolism and/or pERK-PLA2G15 to regulate GPC. However, further studies are needed to address this question. UGDH, UDP-glucose 6-dehydrogenase; SNAI1, Snail family transcriptional repressor 1; PDGFRB, platelet-derived growth factor receptor beta; RELA (NFκB-p65), nuclear factor NF-kappa-B p65 subunit; PPAR $\gamma$ , peroxisome proliferator-activated receptor  $\gamma$ ; pERK, phosphorylated extracellular signal-regulated kinase; PLA2G15, phospholipase A2 group XV.

involved in the EMT programme and likely to participate in normal human breast gland development, that is SERPINE1, RPL26L1, PLOD2, UGDH, LGALS1, VIM, TSTA3, SORD, FDFT1, ATP2A2, MTCH2, PKP3 and JUP, within which, UGDH, TSTA3, SORD and FDFT1 were metabolic enzymes with UGDH possessing the biggest difference between the epithelial and mesenchymal cell lines. UGDH regulated SNAI1, affected cell proliferation and invasion and is associated with patient survival potentially via regulation of the intracellular GPC level. PDGFRB was involved in the regulation of UGDH in mesenchymal cells, likely through NFκB-p65. Further studies on understanding the roles of UGDH on GPC and its relationship with EMT could be valuable in developing novel therapeutics against breast cancer.

**Acknowledgements**

The study received funding from Icelandic centre for research award number: 163254-051; Recipient: Ottar Rolfsson and Doktorstýrki HÁskóla Ísland; Doktorstýrki 2020; Recipient: Qiong Wang.

**Conflict of interest**

The authors declare no conflict of interest.

**Author contributions**

QW conceived the study, carried out the experiments, performed the data analysis, designed the figures and wrote the manuscript. STK performed the metabolomic

untargeted analysis and the flux analysis in the genome-scale metabolic network reconstructions (GEMs); FJ conducted the metabolomic experiments; AIV carried out experiments; LH, DMF and AS performed the proteomic analysis under the supervision of GS. OR conceived, supervised and funded the study, as well as analysed the data and wrote the manuscript. All authors provided critical feedbacks on the manuscript and data analysis.

## Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1002/1878-0261.13172>.

## Data accessibility

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the data set identifier [PXD024164](https://doi.org/10.1002/1878-0261.13172). The mass spectrometry phosphoproteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the data set identifier [PXD025858](https://doi.org/10.1002/1878-0261.13172).

## References

- Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*. 2019;**20**:69–84. <https://doi.org/10.1038/s41580-018-0080-4>
- Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. *Nat Rev Cancer*. 2018;**18**:128–34. <https://doi.org/10.1038/nrc.2017.118>
- Krebs AM, Mitschke J, Lasierra Losada M, Schmalhofer O, Boerries M, Busch H, et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat Cell Biol*. 2017;**19**:518–29. <https://doi.org/10.1038/ncb3513>
- Yang J, Antin P, Bex G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2020;**21**:341–52. <https://doi.org/10.1038/s41580-020-0237-9>
- Aharonov A, Shakked A, Umansky KB, Savidor A, Genzelinakh A, Kain D, et al. ERBB2 drives YAP activation and EMT-like processes during cardiac regeneration. *Nat Cell Biol*. 2020;**22**:1346–56. <https://doi.org/10.1038/s41556-020-00588-4>
- Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, et al. The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol*. 2000;**2**:76–83. <https://doi.org/10.1038/35000025>
- Tam WL, Lu H, Buikhuisen J, Soh BS, Lim E, Reinhardt F, et al. Protein kinase C alpha is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell*. 2013;**24**:347–64. <https://doi.org/10.1016/j.ccr.2013.08.005>
- Yu Y, Luo W, Yang ZJ, Chi JR, Li YR, Ding Y, et al. miR-190 suppresses breast cancer metastasis by regulation of TGF-beta-induced epithelial-mesenchymal transition. *Mol Cancer*. 2018;**17**:70. <https://doi.org/10.1186/s12943-018-0818-9>
- Gudjonsson T, Villadsen R, Nielsen HL, Rønnow-Jessen L, Bissell MJ, Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev*. 2002;**16**:693–706. <https://doi.org/10.1101/gad.952602>
- Sigurðsson V, Hilmarsdóttir B, Sigmundsdóttir H, Fríðriksdóttir AJR, Ringnér M, Villadsen R, et al. Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One*. 2011;**6**:e23833. <https://doi.org/10.1371/journal.pone.0023833>
- Bieric B, Pierce SE, Kroeger C, Stover DG, Pattabiraman DR, Thiru P, et al. Integrin-beta4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proc Natl Acad Sci USA*. 2017;**114**:E2337–46. <https://doi.org/10.1073/pnas.1618298114>
- Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, et al. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev*. 2001;**15**:50–65. <https://doi.org/10.1101/gad.828901>
- Ackland ML, Michalczyk A, Whitehead RH. PMC42, a novel model for the differentiated human breast. *Exp Cell Res*. 2001;**263**:14–22. <https://doi.org/10.1006/excr.2000.5106>
- Cursons J, Leuchowius KJ, Waltham M, Tomaskovic-Crook E, Foroutan M, Bracken CP, et al. Stimulus-dependent differences in signalling regulate epithelial-mesenchymal plasticity and change the effects of drugs in breast cancer cell lines. *Cell Commun Signal*. 2015;**13**:26. <https://doi.org/10.1186/s12964-015-0106-x>
- Whitehead RH, Bertonecello I, Webber LM, Pedersen JS. A new human breast carcinoma cell line (PMC42) with stem cell characteristics. I. Morphologic characterization. *J Natl Cancer Inst*. 1983;**70**:649–61.
- Bhatia S, Monkman J, Blick T, Duijff PH, Nagaraj SH, Thompson EW. Multi-omics characterization of the spontaneous mesenchymal-epithelial transition in the PMC42 breast cancer cell lines. *J Clin Med*. 2019;**8**:1253. <https://doi.org/10.3390/jcm8081253>
- Briem E, Ingthorsson S, Traustadottir GA, Hilmarsdóttir B, Gudjonsson T. Application of the

- D492 cell lines to explore breast morphogenesis, EMT and cancer progression in 3D culture. *J Mammary Gland Biol Neoplasia*. 2019;**24**:139–47. <https://doi.org/10.1007/s10911-018-09424-w>
- 18 Kröger C, Afcyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, et al. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci USA*. 2019;**116**:7353–62. <https://doi.org/10.1073/pnas.1812876116>
  - 19 Morandi A, Taddei ML, Chiarugi P, Giannoni E. Targeting the metabolic reprogramming that controls epithelial-to-mesenchymal transition in aggressive tumors. *Front Oncol*. 2017;**7**:40. <https://doi.org/10.3389/fonc.2017.00040>
  - 20 Li M, Bu X, Cai B, Liang P, Li K, Qu X, et al. Biological role of metabolic reprogramming of cancer cells during epithelial mesenchymal transition (Review). *Oncol Rep*. 2019;**41**:727–41. <https://doi.org/10.3892/or.2018.6882>
  - 21 Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nat Rev Cancer*. 2019;**19**:716–32. <https://doi.org/10.1038/s41568-019-0213-x>
  - 22 Eiriksson FF, Rolfsson O, Ogmundsdottir HM, Haraldsson GG, Thorsteinsdottir M, Halldorsson S. Altered plasmalogen content and fatty acid saturation following epithelial to mesenchymal transition in breast epithelial cell lines. *Int J Biochem Cell Biol*. 2018;**103**:99–104. <https://doi.org/10.1016/j.biocel.2018.08.003>
  - 23 Halldorsson S, Rohatgi N, Magnúsdóttir M, Choudhary KS, Gudjonsson T, Knutsen E, et al. Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett*. 2017;**396**:117–29. <https://doi.org/10.1016/j.canlet.2017.03.019>
  - 24 Karvelsson ST, Sigurdsson A, Seip K, Grinde MT, Wang Q, Johannsson F, et al. EMT-derived alterations in glutamine metabolism sensitize mesenchymal breast cells to mTOR inhibition. *Mol Cancer Res*. 2021;**19**:1546–58. <https://doi.org/10.1158/1541-7786.MCR-20-0962>
  - 25 Karvelsson ST, Wang Q, Hilmarsdóttir B, Sigurdsson A, Moestue SA, Maclandsmo GM, et al. Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer. *NPJ Syst Biol Appl*. 2021;**7**:36. <https://doi.org/10.1038/s41540-021-00195-5>
  - 26 Wang Q, Karvelsson ST, Kotronoulas A, Gudjonsson T, Halldorsson S, Rolfsson O. Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) is upregulated in breast epithelial-mesenchymal transition and responds to oxidative stress. *Mol Cell Proteomics*. 2021, in press. <https://doi.org/10.1016/j.mcpro.2021.100185>
  - 27 Arnold JM, Gu F, Ambati CR, Rasaily U, Ramirez-Pena E, Joseph R, et al. UDP-glucose 6-dehydrogenase regulates hyaluronic acid production and promotes breast cancer progression. *Oncogene*. 2020;**39**:3089–101. <https://doi.org/10.1038/s41388-019-0885-4>
  - 28 Wang X, Liu R, Zhu W, Chu H, Yu H, Wei P, et al. UDP-glucose accelerates SNAIL mRNA decay and impairs lung cancer metastasis. *Nature*. 2019;**571**:127–31. <https://doi.org/10.1038/s41586-019-1340-y>
  - 29 Bontemps Y, Vuillermoz B, Antonicelli F, Perreau C, Danan JL, Maquart FX, et al. Specific protein-1 is a universal regulator of UDP-glucose dehydrogenase expression: its positive involvement in transforming growth factor-beta signaling and inhibition in hypoxia. *J Biol Chem*. 2003;**278**:21566–75. <https://doi.org/10.1074/jbc.M209366200>
  - 30 Clarkin CE, Allen S, Kuiper NJ, Wheeler BT, Wheeler-Jones CP, Pitsillides AA. Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and release, control sulfated GAG synthesis, and promote chondrogenesis. *J Cell Physiol*. 2011;**226**:749–61. <https://doi.org/10.1002/jcp.22393>
  - 31 Fernando R, Smith TJ. Slit2 regulates hyaluronan & cytokine synthesis in fibrocytes: potential relevance to thyroid associated ophthalmopathy. *J Clin Endocrinol Metab*. 2021;**106**:e20–33. <https://doi.org/10.1210/clinem/dgaa684>
  - 32 Lin LH, Chou HC, Chang SJ, Liao EC, Tsai YT, Wei YS, et al. Targeting UDP-glucose dehydrogenase inhibits ovarian cancer growth and metastasis. *J Cell Mol Med*. 2020;**24**:11883–902. <https://doi.org/10.1111/jcmm.15808>
  - 33 Pan YR, Vatsyayan J, Chang YS, Chang HY. Epstein-Barr virus latent membrane protein 2A upregulates UDP-glucose dehydrogenase gene expression via ERK and PI3K/Akt pathway. *Cell Microbiol*. 2008;**10**:2447–60. <https://doi.org/10.1111/j.1462-5822.2008.01221.x>
  - 34 Teoh ST, Ogrodzinski MP, Lunt SY. UDP-glucose 6-dehydrogenase knockout impairs migration and decreases in vivo metastatic ability of breast cancer cells. *Cancer Lett*. 2020;**492**:21–30. <https://doi.org/10.1016/j.canlet.2020.07.031>
  - 35 Goodwin CR, Ahmed AK, Xia S. UDP- $\alpha$ -D-glucose 6-dehydrogenase: a promising target for glioblastoma. *Oncotarget*. 2019;**10**:1542–3. <https://doi.org/10.18632/oncotarget.26670>
  - 36 Huang D, Casale GP, Tian J, Lele SM, Pisarev VM, Simpson MA, et al. Udp-glucose dehydrogenase as a novel field-specific candidate biomarker of prostate cancer. *Int J Cancer*. 2010;**126**:315–27. <https://doi.org/10.1002/ijc.24820>
  - 37 Oyinlade O, Wei S, Lal B, Latorra J, Zhu H, Goodwin CR, et al. Targeting UDP-alpha-D-glucose 6-dehydrogenase inhibits glioblastoma growth and migration. *Oncogene*. 2018;**37**:2615–29. <https://doi.org/10.1038/s41388-018-0138-y>



- 38 Ingthorsson S, Andersen K, Hilmarsdottir B, Maclandsmo GM, Magnusson MK, Gudjonsson T. HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. *Oncogene*. 2016;**35**:4244–55. <https://doi.org/10.1038/onc.2015.489>
- 39 Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003;**75**:663–70. <https://doi.org/10.1021/ac026117i>
- 40 Aass KR, Mjelle R, Kastnes MH, Tryggstad SS, van den Brink LM, Westhlin M, et al. IL-32 is a metabolic regulator promoting survival and proliferation of malignant plasma cells. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.02.22.431638>
- 41 Kil YJ, Becker C, Sandoval W, Goldberg D, Bern M. Preview: a program for surveying shotgun proteomics tandem mass spectrometry data. *Anal Chem*. 2011;**83**:5259–67. <https://doi.org/10.1021/ac200609a>
- 42 Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994;**5**:976–89. [https://doi.org/10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2)
- 43 Horn M, Ueckert T, Fritzsche K, Tham K, Paschke C, Berg F, Pfaff H, Jiang X, Li S, Lopez-Ferrer D. New method for label-free quantification in the Proteome Discoverer framework. 2016.
- 44 Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. 2018;**47**:D442–50. <https://doi.org/10.1093/nar/gky1106>
- 45 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;**26**:1367–72. <https://doi.org/10.1038/nbt.1511>
- 46 Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;**10**:1794–805. <https://doi.org/10.1021/pr101065j>
- 47 Rolfsson O, Johannsson F, Magnusdottir M, Paglia G, Sigurjonsson OE, Bordbar A, et al. Mannose and fructose metabolism in red blood cells during cold storage in SAGM. *Transfusion*. 2017;**57**:2665–76. <https://doi.org/10.1111/trf.14266>
- 48 Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;**78**:779–87. <https://doi.org/10.1021/ac051437y>
- 49 Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*. 2008;**9**:504. <https://doi.org/10.1186/1471-2105-9-504>
- 50 Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;**12**:523–6. <https://doi.org/10.1038/nmeth.3393>
- 51 Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. 2011;**6**:1060–83. <https://doi.org/10.1038/nprot.2011.335>
- 52 De Livera AM, Olshansky G, Simpson JA, Creck DJ. NormalizMets: assessing, selecting and implementing statistical methods for normalizing metabolomics data. *Metabolomics*. 2018;**14**:54. <https://doi.org/10.1007/s11306-018-1347-7>
- 53 Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*. 2002;**18**:S105–10. [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.S105](https://doi.org/10.1093/bioinformatics/18.suppl_1.S105)
- 54 Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat Methods*. 2016;**13**:731–40. <https://doi.org/10.1038/nmeth.3901>
- 55 Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;**37**:1–13. <https://doi.org/10.1093/nar/gkn923>
- 56 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;**4**:44–57. <https://doi.org/10.1038/nprot.2008.211>
- 57 Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2019;**48**:D498–503. <https://doi.org/10.1093/nar/gkz1031>
- 58 Györfi B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*. 2010;**123**:725–31. <https://doi.org/10.1007/s10549-009-0674-9>
- 59 Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: an updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. *J Genet Genomics*. 2019;**46**:595–7. <https://doi.org/10.1016/j.jgg.2019.11.010>

- 60 Loh CY, Chai JY, Tang TF, Wong WF, Sethi G, Shanmugam MK, et al. The E-cadherin and N-cadherin switch in epithelial-to-mesenchymal transition: signaling, therapeutic implications, and challenges. *Cells*. 2019;**8**:1118. <https://doi.org/10.3390/cells8101118>
- 61 Freeze HH, Hart GW, Schnaar RL. Glycosylation precursors. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. editors. *Essentials of glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2015. p. 51–63.
- 62 Li H, Ning S, Ghandi M, Kryukov GV, Gopal S, Deik A, et al. The landscape of cancer cell line metabolism. *Nat Med*. 2019;**25**:850–60. <https://doi.org/10.1038/s41591-019-0404-8>
- 63 Shaul YD, Freinkman E, Comb WC, Cantor JR, Tam WL, Thiru P, et al. Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell*. 2014;**158**:1094–109. <https://doi.org/10.1016/j.cell.2014.07.032>
- 64 Steinhacuser SS, Morera E, Budkova Z, Schepsky A, Wang Q, Rolfsson O, et al. ECM1 secreted by HER2-overexpressing breast cancer cells promotes formation of a vascular niche accelerating cancer cell migration and invasion. *Lab Invest*. 2020;**100**:928–44. <https://doi.org/10.1038/s41374-020-0415-6>
- 65 Barkovskaya A, Goodwin CM, Scip K, Hilmarsdottir B, Petersen S, Stalneckner C, et al. Detection of phenotype-specific therapeutic vulnerabilities in breast cells using a CRISPR loss-of-function screen. *Mol Oncol*. 2021;**15**:2026–45. <https://doi.org/10.1002/1878-0261.12951>
- 66 Aigner K, Descovich L, Mikula M, Sultan A, Dampier B, Bonn e S, et al. The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression. *FEBS Lett*. 2007;**581**:1617–24. <https://doi.org/10.1016/j.febslet.2007.03.026>
- 67 Bacigalupo ML, Manzi M, Espelt MV, Gentilini LD, Compagno D, Laderach DJ, et al. Galectin-1 triggers epithelial-mesenchymal transition in human hepatocellular carcinoma cells. *J Cell Physiol*. 2015;**230**:1298–309. <https://doi.org/10.1002/jcp.24865>
- 68 Bedi U, Mishra VK, Wasilewski D, Scheel C, Johnsen SA. Epigenetic plasticity: a central regulator of epithelial-to-mesenchymal transition in cancer. *Oncotarget*. 2014;**5**:2016–29. <https://doi.org/10.18632/oncotarget.1875>
- 69 Li H, Zhong A, Li S, Meng X, Wang X, Xu F, et al. The integrated pathway of TGF $\beta$ /Snail with TNF $\alpha$ /NF $\kappa$ B may facilitate the tumor-stroma interaction in the EMT process and colorectal cancer prognosis. *Sci Rep*. 2017;**7**:4915. <https://doi.org/10.1038/s41598-017-05280-6>
- 70 Pav n M, Parre o M, T llez-Gabriel M, Sancho F, L pez M, C spedes M, et al. Gene expression signatures and molecular markers associated with clinical outcome in locally advanced head and neck carcinoma. *Carcinogenesis*. 2012;**33**:1707–16. <https://doi.org/10.1093/carcin/bgs207>
- 71 Xu F, Zhang J, Hu G, Liu L, Liang W. Hypoxia and TGF- $\beta$ 1 induced PLOD2 expression improve the migration and invasion of cervical cancer cells by promoting epithelial-to-mesenchymal transition (EMT) and focal adhesion formation. *Cancer Cell Int*. 2017;**17**:54. <https://doi.org/10.1186/s12935-017-0420-z>
- 72 Yamagami Y, Kawami M, Ojima T, Futatsugi S, Yumoto R, Takano M. Role of plasminogen activator inhibitor-1 in methotrexate-induced epithelial-mesenchymal transition in alveolar epithelial A549 cells. *Biochem Biophys Res Comm*. 2020;**525**:543–8. <https://doi.org/10.1016/j.bbrc.2020.02.131>
- 73 Mendez MG, Kojima S-I, Goldman RD. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J*. 2010;**24**:1838–51. <https://doi.org/10.1096/fj.09-151639>
- 74 Hugo H, Ackland ML, Blick T, Lawrence MG, Clements JA, Williams ED, et al. Epithelial-mesenchymal and mesenchymal-epithelial transitions in carcinoma progression. *J Cell Physiol*. 2007;**213**:374–83. <https://doi.org/10.1002/jcp.21223>
- 75 Vergara D, Simeone P, Latorre D, Cascione F, Leporatti S, Trerotola M, et al. Proteomic analysis of E-cadherin knockdown in epithelial breast cancer cells. *J Biotechnol*. 2015;**202**:3–11. <https://doi.org/10.1016/j.jbiotec.2014.10.034>
- 76 Vitale DL, Caon I, Parnigoni A, Sevic I, Spinelli FM, Icardi A, et al. Initial identification of UDP-glucose dehydrogenase as a prognostic marker in breast cancer patients, which facilitates epirubicin resistance and regulates hyaluronan synthesis in MDA-MB-231 cells. *Biomolecules*. 2021;**11**:246.
- 77 Sonkar K, Ayyappan V, Tressler CM, Adelaja O, Cai R, Cheng M, et al. Focus on the glycerophosphocholine pathway in choline phospholipid metabolism of cancer. *NMR Biomed*. 2019;**32**:e4112. <https://doi.org/10.1002/nbm.4112>
- 78 Cao MD, Sitter B, Bathen TF, Bofin A, L nning PE, Lundgren S, et al. Predicting long-term survival and treatment response in breast cancer patients receiving neoadjuvant chemotherapy by MR metabolic profiling. *NMR Biomed*. 2012;**25**:369–78. <https://doi.org/10.1002/nbm.1762>
- 79 Giskoc dg rd GF, Grinde MT, Sitter B, Axelsson DE, Lundgren S, Fjoesne HE, et al. Multivariate modeling and prediction of breast cancer prognostic factors using MR metabolomics. *J Proteome Res*. 2010;**9**:972–9. <https://doi.org/10.1021/pr9008783>
- 80 Moestue SA, Borgan E, Huuse EM, Lindholm EM, Sitter B, B rresen-Dale A-L, et al. Distinct choline

- metabolic profiles are associated with differences in gene expression for basal-like and luminal-like breast cancer xenograft models. *BMC Cancer*. 2010;**10**:433. <https://doi.org/10.1186/1471-2407-10-433>
- 81 Pommier RM, Sanlaville A, Tonon L, Kielbassa J, Thomas E, Ferrari A, et al. Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat Commun*. 2020;**11**:3431. <https://doi.org/10.1038/s41467-020-17249-7>
  - 82 Fougner C, Bergholtz H, Norum JH, Sortie T. Re-definition of claudin-low as a breast cancer phenotype. *Nat Commun*. 2020;**11**:1787. <https://doi.org/10.1038/s41467-020-15574-5>
  - 83 Bharti SK, Mironchik Y, Wildes F, Penet MF, Goggins E, Krishnamachary B, et al. Metabolic consequences of HIF silencing in a triple negative human breast cancer xenograft. *Oncotarget*. 2018;**9**:15326–39. <https://doi.org/10.18632/oncotarget.24569>
  - 84 Koch K, Hartmann R, Schröter F, Suwala AK, Maciaczyk D, Krüger AC, et al. Reciprocal regulation of the cholinergic phenotype and epithelial-mesenchymal transition in glioblastoma cells. *Oncotarget*. 2016;**7**:73414–31. <https://doi.org/10.18632/oncotarget.12337>
  - 85 Okazaki Y, Nakamura K, Takeda S, Yoshizawa I, Yoshida F, Ohshima N, et al. GDE5 inhibition accumulates intracellular glycerophosphocholine and suppresses adipogenesis at a mitotic clonal expansion stage. *Am J Physiol Cell Physiol*. 2019;**316**:C162–74. <https://doi.org/10.1152/ajpcell.00305.2018>
  - 86 Ahmadian M, Suh JM, Hah N, Liddle C, Atkins AR, Downes M, et al. PPARgamma signaling and metabolism: the good, the bad and the future. *Nat Med*. 2013;**19**:557–66. <https://doi.org/10.1038/nm.3159>
  - 87 Tontonoz P, Spiegelman BM. Fat and beyond: the diverse biology of PPARgamma. *Annu Rev Biochem*. 2008;**77**:289–312. <https://doi.org/10.1146/annurev.biochem.77.061307.091829>
  - 88 Iorio E, Mezzanzanica D, Alberti P, Spadaro F, Ramoni C, D'Ascenzo S, et al. Alterations of choline phospholipid metabolism in ovarian tumor progression. *Cancer Res*. 2005;**65**:9369–76. <https://doi.org/10.1158/0008-5472.CAN-05-1146>
  - 89 Menzel N, Fischl W, Hueging K, Bankwitz D, Frentzen A, Haid S, et al. MAP-kinase regulated cytosolic phospholipase A2 activity is essential for production of infectious hepatitis C virus particles. *PLoS Pathog*. 2012;**8**:e1002829. <https://doi.org/10.1371/journal.ppat.1002829>
  - 90 Ulisse S, Cinque B, Silvano G, Rucci N, Biordi L, Cifone MG, et al. Erk-dependent cytosolic phospholipase A2 activity is induced by CD95 ligand cross-linking in the mouse derived sertoli cell line TM4 and is required to trigger apoptosis in CD95 bearing cells. *Cell Death Differ*. 2000;**7**:916–24. <https://doi.org/10.1038/sj.cdd.4400716>
  - 91 Huber MA, Azoitei N, Baumann B, Grunert S, Sommer A, Pehamberger H, et al. NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest*. 2004;**114**:569–81. <https://doi.org/10.1172/JCI21358>
  - 92 Naidu S, Shi L, Magesc P, Middleton JD, Laganá A, Sahoo S, et al. PDGFR-modulated miR-23b cluster and miR-125a-5p suppress lung tumorigenesis by targeting multiple components of KRAS and NF-kB pathways. *Sci Rep*. 2017;**7**:15441. <https://doi.org/10.1038/s41598-017-14843-6>
  - 93 Wang H, Yin Y, Li W, Zhao X, Yu Y, Zhu J, et al. Over-expression of PDGFR-beta promotes PDGF-induced proliferation, migration, and angiogenesis of EPCs through PI3K/Akt signaling pathway. *PLoS One*. 2012;**7**:e30503. <https://doi.org/10.1371/journal.pone.0030503>
  - 94 Wang Z, Ahmad A, Li Y, Kong D, Azmi AS, Banerjee S, et al. Emerging roles of PDGF-D signaling pathway in tumor development and progression. *Biochem Biophys Acta*. 2010;**1806**:122–30. <https://doi.org/10.1016/j.bbcan.2010.04.003>
  - 95 Wu Q, Wang R, Yang Q, Hou X, Chen S, Hou Y, et al. Chemoresistance to gemcitabine in hepatoma cells induces epithelial-mesenchymal transition and involves activation of PDGF-D pathway. *Oncotarget*. 2013;**4**:1999–2009.
  - 96 Shimamura T, Hsu TC, Colburn NH, Bejcek BE. Activation of NF-kappaB is required for PDGF-B chain to transform NIH3T3 cells. *Exp Cell Res*. 2002;**274**:157–67. <https://doi.org/10.1006/excr.2001.5449>
  - 97 Xu B, Bai Z, Yin J, Zhang Z. Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. *PeerJ*. 2019;**7**:e7091. <https://doi.org/10.7717/peerj.7091>
  - 98 Piskareva O, Harvey H, Nolan J, Conlon R, Alcock L, Buckley P, et al. The development of cisplatin resistance in neuroblastoma is accompanied by epithelial to mesenchymal transition in vitro. *Cancer Lett*. 2015;**364**:142–55. <https://doi.org/10.1016/j.canlet.2015.05.004>
  - 99 Ren B-J, Zhou Z-W, Zhu D-J, Ju Y-L, Wu J-H, Ouyang M-Z, et al. Alisertib induces cell cycle arrest, apoptosis, autophagy and suppresses EMT in HT29 and Caco-2 cells. *Int J Mol Sci*. 2015;**17**:41. <https://doi.org/10.3390/ijms17010041>
  - 100 Wang Y, Zeng G, Jiang Y. The emerging roles of miR-125b in cancers. *Cancer Manag Res*. 2020;**12**:1079–88. <https://doi.org/10.2147/cmar.S232388>

- 101 Schwab A, Siddiqui A, Vazakidou ME, Napoli F, Böttcher M, Menchicchi B, et al. Polyol pathway links glucose metabolism to the aggressiveness of cancer cells. *Cancer Res.* 2018;**78**:1604–18. <https://doi.org/10.1158/0008-5472.Can-17-2834>
- 102 Yang Y-F, Chang Y-C, Jan Y-H, Yang C-J, Huang M-S, Hsiao M. Squalene synthase promotes the invasion of lung cancer cells via the osteopontin/ERK pathway. *Oncogenesis.* 2020;**9**:78. <https://doi.org/10.1038/s41389-020-00262-2>
- 103 Davis FM, Parsonage MT, Cabot PJ, Parat M-O, Thompson EW, Roberts-Thomson SJ, et al. Assessment of gene expression of intracellular calcium channels, pumps and exchangers with epidermal growth factor-induced epithelial-mesenchymal transition in a breast cancer cell line. *Cancer Cell Int.* 2013;**13**:76. <https://doi.org/10.1186/1475-2867-13-76>
- 104 Demirkan B. The roles of epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET) in breast cancer bone metastasis: potential targets for prevention and treatment. *J Clin Med.* 2013;**2**:264–82. <https://doi.org/10.3390/jcm2040264>
- 105 Lourenco AR, Ban Y, Crowley MJ, Lee SB, Ramchandani D, Du W, et al. Differential contributions of pre- and post-EMT tumor cells in breast cancer metastasis. *Can Res.* 2020;**80**:163–9. <https://doi.org/10.1158/0008-5472.Can-19-1427>

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Study workflow and the phenotypes of the cell lines. (A) Workflow of the proteomic analysis of the three breast EMT cell models and metabolomics analysis after siRNA knock-down of the metabolic target *UGDH* in all the mesenchymal cell lines. Three breast EMT cell models (epithelial and mesenchymal cell line pairs) were used in this study, D492&D492M, HMLE&HMLEM, and PMC42LA&PMC42ET. The proteomic strategy was label-free quantification (LFQ) with each cell line in triplicates. The metabolomic strategy was untargeted metabolomics in negative, positive, and basic modes with six replicates. The upstream signaling regulation and downstream cellular functions of UGDH were also investigated in this study. The tumorigenic breast mesenchymal cell line D492HER2 and malignant MDA-MB-231 were employed further to define the functions of UGDH in tumor malignancy. (B) Photos of all the cell lines in the three breast EMT cell models used in this study were shown. Different cell lines were cultured in their

routine maintaining medium respectively, and the photos were taken under phase contrast with objectives 5x or 20x.

**Fig. S2.** Inconsistent EMT markers. A list of known EMT markers (based on the public EMT database dbEMT) was inconsistently altered among the three EMT models. Student's T-test, \*:  $P < 0.05$ ; \*\*:  $P < 0.01$ ; \*\*\*:  $P < 0.001$ ;  $n = 3$ . CD44, CD44 antigen; LMNB1, Lamin-B1; MSN, Moesin; FLNA, Filamin-A; TLN1, Talin-1; FSCN1, Fascin; EGFR, Epidermal growth factor receptor; S100A2, S100 calcium binding protein A2; NDRG1, N-myc downstream regulated 1.

**Fig. S3.** Accuracy and validity of the proteomic analysis. The accuracy and validity of the proteomic analysis in this study were confirmed by comparing the current data to our previously generated proteomic data for the D492 EMT model [26]. The correlation between these two datasets was 0.936. The high correlation coefficient (Pearson correlation, 0.936) of the datasets ensures good accuracy and validity of the proteomic analysis in this study. It laid the foundation for valid conclusions deduced from this study.

**Fig. S4.** Functional annotation of the GO terms (CC and MF) for the three EMT models. Functional annotation of the GO terms (CC and MF) was conducted on the DAVID (DAVID Bioinformatics Resources 6.8) platform for each EMT model. Data used for the GO annotation analysis (Supplementary Table 3) were proteins significantly altered in each EMT model (Permutation-based FDR < 0.05). Default settings were used for the analysis. The GO terms were listed according to the  $-\log_{10}$  p value in descending order. The numbers of genes in each GO term were also plotted as dots/line plots. CC: Cellular Component; MF: Molecular Function.

**Fig. S5.** Knock-down efficiency of *UGDH* with two siRNAs. (A–D) The knock-down efficiency of *UGDH* with two siRNAs compared to the scramble control was around 80 % in D492M ( $n = 7$  for the first siRNA;  $n = 9$  for the second siRNA) (A–B) and 60 % in D492HER2 ( $n = 5$ ) (C–D). (E) The knock-down efficiency of *UGDH* with two siRNAs in the metabolomics experiments for D492M, HMLEM, and PMC42ET was 90 % ( $n = 5$ ). KD: Knock-down. Student's T-test, \*\*\*:  $P < 0.001$ . UGDH, UDP-glucose 6-dehydrogenase.

**Fig. S6.** Functional analysis of UGDH in EMT. (A) Photos of the D492M and D492HER2 cells following knock-down of *UGDH* via two siRNAs in the invasion assay. Cells were stained with DAPI and observed under the objective 10x. (B–C) Cell invasion decreased with the second siRNA knock-down of *UGDH* in both non-tumorigenic D492M (B) and tumorigenic

D492HER2 (C). *n* equals 3, and ten spots were chosen for each replicate during the cell counting process. (D-E) One of the main EMT transcription factors SNAIL was downregulated following the second siRNA knock-down of *UGDH* in both non-tumorigenic D492M (*n* = 5) (D) and tumorigenic D492HER2 (*n* = 4) (E). (F-H) The Kaplan-Meier plots of FDFT1, SORD, and TSTA3 in basal breast cancer patients were downloaded from kmplot.com. Student's T-test, \*\*:  $P < 0.01$ ; \*\*\*:  $P < 0.001$ . UGDH, UDP-glucose 6-dehydrogenase; SNAIL, Snail Family Transcriptional Repressor 1; FDFT1, Squalene synthase; SORD, Sorbitol dehydrogenase; TSTA3, GDP-L-fucose synthase.

**Fig. S7.** GPC and NAA were altered with the siUGDH treatment in D492HER2 and MDA-MB-231. (A-D) The glycerophosphocholine (GPC) level was decreased, and the acetylaspartate (NAA) level was increased after the siUGDH treatment in the tumorigenic D492HER2 (*n* = 3) and malignant MDA-MB-231 (*n* = 6) cell lines. (E) There were no significant differences to the GPC levels between mesenchymal cells and non-mesenchymal cells based on published datasets in literature [62, 63] (Supplementary Table 5). (F-G) The expression levels of choline (F) and phosphocholine (G) with *UGDH* knock-down in the three EMT cell models. No significant and consistent changes were observed for both metabolites in all the cell lines (*n* = 5). (H) UGDH has been reported to downregulate PPAR $\gamma$  [27]. To test if there was a negative correlation between UGDH and PPAR signaling, we performed a phosphoproteomic analysis on the D492 EMT cell model (Supplementary Table 7) and noticed that the PPAR signaling was downregulated in the mesenchymal cells where UGDH was highly expressed. The IPA pathways were listed based on the  $-\log_{10}(p \text{ value})$ , and the *z*-scores for the pathways were represented by the dots/line plot. Red: higher in the mesenchymal D492M; blue: higher in the epithelial D492. (I) The enzyme PLA2G15 potentially involved in the hydrolysis of phosphatidylcholine (PtdCho) into GPC was higher in both D492M and D492HER2. UGDH has been reported to regulate the phosphorylation of ERK (pERK) [32]. cPLA2 is responsible for GPC synthesis from PtdCho in choline metabolism

and is under the control of ERK/MAPK [89, 90]. We also observed that PLA2G15 was highly expressed in D492M and D492HER2 compared to D492 (I), suggesting the knock-down of *UGDH* may downregulate GPC via pERK-PLA2G15 (*n* = 3). Student's T-test, \*:  $P < 0.05$ ; \*\*:  $P < 0.01$ ; \*\*\*:  $P < 0.001$ . n.s.: not significant. UGDH, UDP-glucose 6-dehydrogenase; PLA2G15, Phospholipase A2 group XV. **Fig. S8.** PDGFRB regulates UGDH via RELA (NF $\kappa$ B-p65). (A) PDGFRB was highly expressed in the tumorigenic mesenchymal cell line D492HER2 based on the RPPA analysis (*n* = 3) [65]. (B) The knock-down efficiency of *PDGFRB* with siRNA in the D492HER2 cell line was about 90 % (*n* = 6). (C) RELA (NF $\kappa$ B-p65) was downregulated after the siRNA knock-down of *PDGFRB* in D492HER2 (*n* = 6). (D) UGDH was downregulated after the siRNA knock-down of *PDGFRB* in D492HER2 (*n* = 6). (E) The knock-down efficiency of *RELA* with the second siRNA in D492M was around 70 % (*n* = 6). (F) UGDH was downregulated after the knock-down of *RELA* in D492M with the second siRNA (*n* = 6). (G) The knock-down efficiency of *RELA* with the first siRNA in D492HER2 was around 90 % (*n* = 6). (H) No significant change in UGDH was observed after the knock-down of *RELA* with the first siRNA in D492HER2 (*n* = 6). (I) The knock-down efficiency of *RELA* with the second siRNA in the D492HER2 cell line was about 90 % (*n* = 6). (J) UGDH was downregulated after the knock-down of *RELA* in D492HER2 with the second siRNA (*n* = 6). Student's T-test, \*:  $P < 0.05$ ; \*\*:  $P < 0.01$ ; \*\*\*:  $P < 0.001$ . UGDH, UDP-glucose 6-dehydrogenase; PDGFRB, Platelet-derived growth factor receptor beta; RELA (NF $\kappa$ B-p65), Nuclear factor NF-kappa-B p65 subunit.

**Table S1.** The internal standard mix used in the metabolomics analysis.

**Table S2.** A list of primers used in this study.

**Table S3.** Perseus output data.

**Table S4.** Raw data of proteomics.

**Table S5.** Publicly available data on the GPC levels of mesenchymal cells.

**Table S6.** *In silico* knockdown of UGDH in GEMs.

**Table S7.** Data of phosphoproteomics.



## Paper II





# Glutamine-Fructose-6-Phosphate Transaminase 2 (GFPT2) Is Upregulated in Breast Epithelial–Mesenchymal Transition and Responds to Oxidative Stress

## Authors

Qiong Wang, Sigurdur Trausti Karvelsson, Aristotelis Kotronoulas, Thorarinn Gudjonsson, Skarphedinn Halldorsson, and Ottar Rolfsson

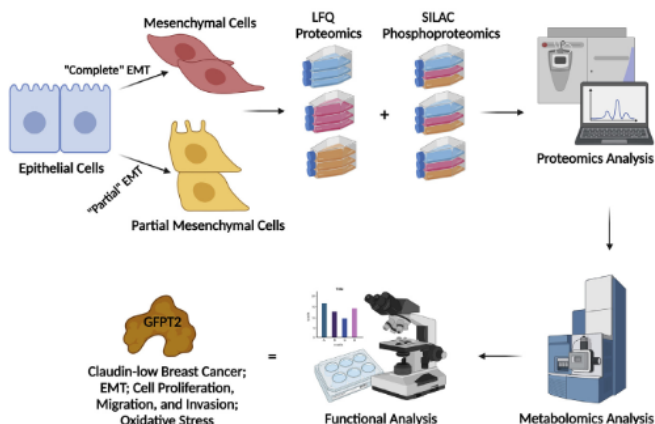
## Correspondence

ottarr@hi.is

## Graphical Abstract

### In Brief

Epithelial–mesenchymal transition (EMT) is a cellular process inherent to cancer cell metastasis. Metabolic reprogramming is a driver of EMT. We performed proteomic profiling of three isogenic cell lines from human breast epithelium representing the epithelial, mesenchymal, and “partial” mesenchymal states of EMT to identify metabolic vulnerabilities associated with cell invasion. Bioinformatic and functional analysis revealed that the metabolic enzyme GFPT2 is a marker of claudin-low breast cancer, responds to oxidative stress, and impacts EMT, cell growth, and cell invasion.



### Highlights

- GFPT2 is upregulated following EMT.
- GFPT2 is a marker for claudin-low breast cancer.
- GFPT2 affects vimentin, cell proliferation, and cell invasion.
- GFPT2 responds to oxidative stress.
- GFPT2 is regulated by insulin and EGF.

2022, Mol Cell Proteomics 21(2), 100185

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1016/j.mcpro.2021.100185>



# Glutamine-Fructose-6-Phosphate Transaminase 2 (GFPT2) Is Upregulated in Breast Epithelial–Mesenchymal Transition and Responds to Oxidative Stress

Qiong Wang<sup>1</sup>, Sigurdur Trausti Karvelsson<sup>1</sup>, Aristotelis Kotronoulas<sup>1</sup>, Thorarinn Gudjonsson<sup>2</sup>, Skarphedinn Halldorsson<sup>1</sup>, and Ottar Rolfsson<sup>1,\*</sup>

Breast cancer cells that have undergone partial epithelial–mesenchymal transition (EMT) are believed to be more invasive than cells that have completed EMT. To study metabolic reprogramming in different mesenchymal states, we analyzed protein expression following EMT in the breast epithelial cell model D492 with single-shot LFQ supported by a SILAC proteomics approach. The D492 EMT cell model contains three cell lines: the epithelial D492 cells, the mesenchymal D492M cells, and a partial mesenchymal, tumorigenic variant of D492 that over-expresses the oncogene HER2. The analysis classified the D492 and D492M cells as basal-like and D492HER2 as claudin-low. Comparative analysis of D492 and D492M to tumorigenic D492HER2 differentiated metabolic markers of migration from those of invasion. Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) was one of the top dysregulated enzymes in D492HER2. Gene expression analysis of the cancer genome atlas showed that GFPT2 expression was a characteristic of claudin-low breast cancer. siRNA-mediated knockdown of *GFPT2* influenced the EMT marker vimentin and both cell growth and invasion *in vitro* and was accompanied by lowered metabolic flux through the hexosamine biosynthesis pathway (HBP). Knockdown of *GFPT2* decreased cystathionine and sulfide:quinone oxidoreductase (SQOR) in the transsulfuration pathway that regulates H<sub>2</sub>S production and mitochondrial homeostasis. Moreover, GFPT2 was within the regulation network of insulin and EGF, and its expression was regulated by reduced glutathione (GSH) and suppressed by the oxidative stress regulator GSK3-β. Our results demonstrate that GFPT2 controls growth and invasion in the D492 EMT model, is a marker for oxidative stress, and associated with poor prognosis in claudin-low breast cancer.

Breast cancer is the most prevalent cancer in women worldwide (1). Within 3 years after the initial diagnosis, around 10 to 15% of patients with breast cancer develop distant metastasis (2). Epithelial–mesenchymal transition (EMT) is a natural process during embryonic development that tumor cells hijack to gain migration and invasive properties (3, 4). EMT is characterized by a broad spectrum of epithelial–mesenchymal states that ultimately affect cancer malignancy (5, 6).

Multiple changes to metabolism accompany breast cancer. These include changes to enzymes in glycolysis (7), the tricarboxylic acid (TCA) cycle (8, 9), and fatty acid synthesis (10). More recently, changes in serine and proline biosynthesis (8, 11) and nucleotide metabolism (12) have been described. However, definitive metabolic phenotypes that differentiate between noninvasive complete EMT and partial EMT with invasive potentials remain elusive (13). Understanding how regulation of enzyme activity on the protein level affects invasiveness may improve breast cancer personalized therapeutic interventions.

In this study, we set out to define changes in metabolic enzymes that accompany EMT in the EMT cell model D492 (14, 15) reviewed in Briem *et al.*, 2019 (16). The D492 breast EMT cell model contains three isogenic phenotypes: the epithelial D492 cells, the mesenchymal D492M cells, and the partial mesenchymal D492HER2 cells. D492 is a basal-like human breast epithelial cell line derived from normal tissue. The D492 cell line expresses both luminal (K8, K19) and myoepithelial (K5/6, K14) cytokeratins. It has epithelial stem cell properties and can differentiate into luminal and myoepithelial cells (15, 17). EMT in cultured breast epithelial cells can

From the <sup>1</sup>Center for Systems Biology, Biomedical Center, and <sup>2</sup>Stem Cell Research Unit, Biomedical Center, Department of Anatomy, Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland

\*For correspondence: Ottar Rolfsson, [ottarr@hi.is](mailto:ottarr@hi.is).

Present address for Skarphedinn Halldorsson: Institute for Surgical Research, Vilhelm Magnus Laboratory, Oslo University Hospital, Oslo, Norway.

be triggered with growth factors such as TGF- $\beta$  and EGF (14, 18, 19). It can also be induced via overexpression of certain EMT markers such as TWIST (20). D492M is a mesenchymal cell line spontaneously generated by 3D coculture of D492 with human endothelial cells in the absence of any dominant EMT inducers (15). Although the D492 and D492M cells are premalignant and not tumorigenic, the D492 cells gain tumorigenicity when HER2 is overexpressed. The D492HER2 cells have a partial mesenchymal phenotype, indicating that cells have gone through EMT (14). The D492 EMT cell model thus comprises three cell lines allowing different states of EMT to be studied *in vitro*. We hypothesized that comparative proteomics analysis of these three cell lines would highlight crucial metabolic enzymes to EMT in breast epithelium and discriminate metabolic enzymes that impart invasion properties.

We have previously defined changes to metabolism in the D492 EMT model within genome-scale metabolic network models. Glycan metabolism, amongst others, was altered in EMT (9). These models were based on changes to gene expression and extracellular metabolomic measurements. In this study, we analyzed the metabolic changes in EMT on the protein level, emphasizing mesenchymal cells that possess invasive potentials. We first positioned the D492 EMT cell model within the breast cancer cell model landscape based upon the LFQ and SILAC proteomics data. We then identified the hexosamine biosynthesis pathway (HBP) rate-limiting enzyme, glutamine-fructose-6-phosphate transaminase 2 (GFPT2), as a potential target in claudin-low breast cancer progression.

Enzymes involved in glycan processing were over-represented in both mesenchymal proteomes, and the HBP rate-limiting enzyme GFPT2 was upregulated in D492M and further still in the D492HER2 mesenchymal cells as compared with D492. Metabolomics analysis confirmed changes to HBP flux. We then compared GFPT2 expression across clinical breast cancer subtypes and breast cancer cell lines and knocked down *GFPT2* to assess its effects on the EMT program, cell growth, and cell invasion in the D492 EMT model. These analyses suggest that GFPT2 is a tumor promoter in claudin-low breast cancer. The role of GFPT2 in mediating glycan synthesis has been reported in a series of studies that show that GFPT2 mediates response via glycosylation of master regulators of metabolism, including NF- $\kappa$ B and  $\beta$ -catenin (21, 22). The function of GFPT2 in glutaminolysis is less explored, although its importance has been inferred from its enzymatic activity. In light of recent results that show that altered glutaminolysis in the D492 EMT model influences their ability to synthesize glutathione from glutamine-derived glutamate, and that this influences their susceptibility to cancer therapeutics (23), we explored the role of GFPT2 in maintaining redox balance in EMT.

## EXPERIMENTAL PROCEDURES

### Cell Culture

D492, D492M, D492HER2, and D492DEE were generated in-house (14, 15, 17) and cultured in serum-free H14 medium. The MDA-MB-231 cell line was cultured in RPMI 1640 (Thermo, 52400-025) supplemented with 10% Fetal Bovine Serum (FBS, Gibco 10270106) and 100 IU penicillin and 0.1 mg/ml streptomycin (Gibco, 15140122). Cells were at 37 °C and 5% CO<sub>2</sub> for routine maintenance. The H14 medium is Dulbecco's modified Eagle's medium - F12 (DMEM/F12 without glutamine, Thermo, 21331020) supplemented with 250 ng/ml insulin (Merck, I6634), 10  $\mu$ g/ml transferrin (Merck, T2252), 10 ng/ml EGF (PeproTech, AF-100-15), 2.6 ng/ml Na-selenite (BD Biosciences, 354201), 10<sup>-10</sup> M estradiol (Sigma, E2758), 1.4  $\times$  10<sup>-6</sup> M hydrocortisone (Sigma, H0888), 0.15 IU prolactin (PeproTech, 100-07), 100 IU penicillin & 0.1 mg/ml streptomycin, and 2 mM glutamine (Thermo, 25030024). In the SILAC proteomic experiment, DMEM-F12 was replaced by "DMEM-F-12 for SILAC" (Thermo, 88370) with "light-," "medium-," or "heavy-labeled" arginine or lysine (Cambridge Isotope Laboratories). In the <sup>13</sup>C labeling experiment, the base medium was changed to "DMEM, no glucose, no glutamine, no phenol red" (Thermo, A1443001), and <sup>13</sup>C labeled 1,2-glucose, <sup>13</sup>C labeled L-glutamine, or <sup>13</sup>C labeled 5-glutamine from Cambridge Isotope Laboratories was added. Medium excluded penicillin and streptomycin for the transient knockdown experiments according to instruction. In the invasion assay, H14 was supplemented with 10% FBS. Cell cultures were routinely checked for *mycoplasma* contamination.

### Label-free Quantification (LFQ) Proteomics

**Protein and Peptide Sample Preparation**—Cells were cultured in T75 flasks in triplicates (three flasks per cell line), and the seeding density was 600,000 cells per flask. Proteins were harvested at 90% confluency, and 72 h after seeding, cells were washed twice with ice-cold PBS and lysed by 450  $\mu$ l lysis buffer containing 4% sodium dodecyl sulfate (SDS, MP Biomedicals) in 100 mM Tris (Sigma). Flasks were kept on ice for 10 min. The cell lysates were transferred to 1.5 ml Eppendorf tubes. After five freeze (–80 °C)/thaw (room temperature) cycles, the sample was spun at 20,718g for 20 min at 4 °C. The supernatant was collected and aliquoted in new tubes and stored at –80 °C. Protein quantification was measured with BCA protein assay (Pierce).

For Filter-Aided Sample Preparation (FASP), an equivalent of 300  $\mu$ g of proteins in 150  $\mu$ l from each sample was reduced with 100 mM dithiothreitol (DTT), and samples were then processed using FASP protocol (24). Proteins on the filters were digested twice at 30 °C with trypsin (enzyme-to-substrate ratio: 1:100 (w/w); 3  $\mu$ g  $\times$  2), first overnight and then for another 6 h in a final volume of 200  $\mu$ l. The resulting peptides were desalted using a C18 solid-phase extraction cartridge (Empore, Agilent technologies). Peptides were resuspended in 50  $\mu$ l 1% formic acid and quantified using pierce quantitative colorimetric peptide assay (product 23275, Thermo Scientific).

**LC-MS/MS Analysis**—Trypsin-digested peptides were separated using an Ultimate 3000 RSLC (Thermo Scientific) nanoflow LC system. In total, 130 ng of peptides was loaded with a constant flow of 5  $\mu$ l/min onto an Acclaim PepMap100 nanoViper C18 trap column (100  $\mu$ m inner-diameter, length: 2 cm; Thermo Scientific). After trap enrichment, peptides were eluted onto an EASY-Spray PepMap RSLC nanoViper, C18, particle size: 2  $\mu$ m, pore size: 100 Å column (75  $\mu$ m inner-diameter, length: 50 cm; Thermo Scientific) with a linear gradient of 2 to 35% solvent B (80% acetonitrile with 0.08% formic acid, Solvent A – 0.1% formic acid) over 124 min with a constant flow of 300 nL/min and column temperature of 50 °C. The HPLC system was coupled to a linear ion trap Orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos,



Thermo Scientific) via an EASY-Spray ion source (Thermo Scientific). The spray voltage was set to 1.8 kV, and the temperature of the heated capillary was set to 250 °C. Full-scan MS survey spectra ( $m/z$  335–1800) in profile mode were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 1,000,000 ions. The 15 most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision-induced dissociation (CID, normalized collision energy, 35%; activation Q, 0.250; and activation time, 10 ms) in the LTQ after the accumulation of 5000 ions. Maximal filling times were 1000 ms for the full scans and 150 ms for the MS/MS scans. Precursor ion charge state screening was enabled, and all unassigned charge states, as well as singly charged species, were rejected. The lock mass option was enabled for survey scans to improve mass accuracy (25). Data were acquired using the Xcalibur software.

**Peptide and Protein Identification and Quantification**—The raw mass spectrometric data files were collated into a single quantitated dataset using MaxQuant (version 1.5.2.8) (26) and the Andromeda search engine software (27). Enzyme specificity was set to that of trypsin, allowing for cleavage N-terminal to proline residues and between aspartic acid and proline residues. Other parameters used were: (i) variable modifications—methionine oxidation, protein N-acetylation,  $\text{gln} \rightarrow \text{pyro-glu}$ , phospho (STY), deamidation (NQ); (ii) fixed modifications, cysteine carbamidomethylation; (iii) database: Uniprot-human-up5640 (release date of sequence database searched: 05.2017; number of entries: 20,201); (iv) LFC: min ratio count, 2 (v) MS/MS tolerance: FTMS- 10 ppm, ITMS- 0.6 Da; (vi) maximum peptide length, 6; (vii) maximum missed cleavages, 2; (viii) maximum labeled amino acids, 3; and (ix) false discovery rate (FDR), 1%. LFC intensities were reported individually for each sample and were given as a relative protein quantitation across all samples. LFC intensities were represented by a normalized intensity profile as described by Cox (28) affording a matrix with number of samples and number of protein groups as dimensions. The iBAQ quantification was carried out in MaxQuant (version 1.5.2.8) for the same raw data obtained. The same parameters as described above for the LFC quantification were applied for the iBAQ quantification except for the selection of the iBAQ method for outputs.

Protein identification was defined as one or more identified peptides observed in at least two out of three replicates in at least one cell line. Protein quantification was calculated when at least two out of three replicates in at least one cell line had detectable intensities.

#### Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) (Phospho)Proteomics

**Protein and Peptide Sample Preparation, Fractionation, and Enrichment—Protein Extraction.** The cell lines D492M, D492, and D492HER2 were labeled with “light,” “medium,” and “heavy” stable isotope-labeled versions of arginine and lysine for SILAC analysis, respectively. The SILAC labeling was not randomized among the cell lines for the triplicates. Cells were first cultured in T25 flasks with respective SILAC labels to get fully labeled cell populations for D492 (“medium” label, L-arginine- $^{13}\text{C}_6$  hydrochloride (Arg +6 Da), L-lysine-4,4,5,5-d4 hydrochloride (Lys +4 Da)), D492M (“light” label, L-arginine, L-lysine), and D492HER2 (“heavy” label, L-arginine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$  hydrochloride (Arg +10 Da), L-lysine- $^{13}\text{C}_6$ ,  $^{15}\text{N}_2$  hydrochloride (Lys +8 Da)). The D492 and D492M cells were cultured in the “medium-” and “light-labeled” medium for six passages to ensure that the cells were close to the fully labeled status. The D492HER2 cells were cultured in the “heavy-labeled” medium for five passages. To harvest enough proteins, cells were propagated in T75 flasks (Santa cruz), then cultured in T182 flasks (Santa cruz) in triplicates, and the seeding density was 1,500,000 cells per flask, which was calculated to be consistent with the LFC proteomics experiment. The same procedures as described in the LFC protein preparation section were conducted for SILAC protein

extraction with lysis buffer supplemented with one tablet of PhosSTOP phosphatase inhibitors (Roche) and one tablet of cComplete mini EDTA-free protease inhibitors (Roche).

**Protein Digestion (FASP Processing of Samples).** Proteins were solubilized in 150  $\mu\text{l}$  of Tris-HCl (100 mM, pH 7.6) containing 4% SDS and 100 mM DTT. Protein extracts were heated at 95 °C, and DNA was shredded by sonication to reduce the viscosity of the lysates. Samples were then centrifuged and processed using FASP protocol (24) with some modifications. After lysates were passed through the filters (Nanosep, 10k, PALL Life Sciences), proteins were alkylated in 100  $\mu\text{l}$  iodoacetamide (IAA) at a final concentration of 50 mM for 15 min, filters were washed four times with 200  $\mu\text{l}$  8 M urea in Tris-HCl (100 mM, pH 8), then twice with 200  $\mu\text{l}$  40 mM ammonium bicarbonate. Proteins on the filters were then digested twice at 30 °C with trypsin (enzyme-to-substrate ratio: 1:100 (w/w); 3.3  $\mu\text{g} \times 2$ ), first overnight and then for another 6 h in 200  $\mu\text{l}$ , ammonium bicarbonate at 40 mM. The resulting tryptic peptides were desalted using a C18 solid-phase extraction cartridge (Empore, Agilent technologies).

**Peptide Fractionation (High pH Reverse-phase Fractionation).** Samples equivalent to 4 mg were dissolved in 200  $\mu\text{l}$  of 10 mM ammonium formate buffer (pH 9.5), and peptides were fractionated using high pH RP chromatography. A C18 column from Waters (XBridge peptide BEH, pore size: 130 Å, particle size: 3.5  $\mu\text{m}$ , inner-diameter: 4.6  $\times$  length: 150 mm, Ireland) with a guard column (XBridge, C18, particle size: 3.5  $\mu\text{m}$ , inner-diameter: 4.6  $\times$  length: 20 mm, Waters) was used on an Ultimate 3000 HPLC (Thermo-Scientific). Buffers A and B used for fractionation consisted, respectively of 10 mM ammonium formate in distilled, deionized water (Buffer A) and 10 mM ammonium formate in 90% acetonitrile (Buffer B), and both buffers were adjusted to pH 9.5 with ammonia. Fractions were collected using a WPS-3000FC autosampler (Thermo-Scientific) at 1 min intervals. Column and guard column were equilibrated with 2% buffer B for 20 min at a constant flow rate of 0.75 ml/min and a constant temperature of 21 °C. Samples (185  $\mu\text{l}$ ) were loaded onto the column at 0.75 ml/min, and the separation gradient started from 2% buffer B to 5% B in 6 min, then from 5% B to 60% B within 55 min. The column was washed for 7 min at 100% buffer B and equilibrated at 2% buffer B for 20 min, as mentioned above. The fraction collection started 1 min after injection and stopped after 80 min (total of 80 fractions, 750  $\mu\text{l}$  each). Each peptide fraction was acidified immediately after elution from the column by adding 20 to 30  $\mu\text{l}$  10% formic acid to each tube in the autosampler. The total number of fractions concatenated was set to 10, with 96% of material from each fraction was used for phospho-enrichment, and 4% was used for total proteome analysis. The content of the fraction from each set was dried prior to further analysis.

**Phosphoproteomic Phospho-peptide Enrichment.** Phospho-peptide enrichment was performed using MagReSyn-TiIMAC beads (Resyn Biosciences) and Magnetic Rack (DynaMag-2, Life Technologies). Tryptic peptides to TiIMAC beads were used at 1:5 ratio (w/w). Beads were first washed using Magnetic Rack with 80  $\mu\text{l}$ , 1%  $\text{NH}_4\text{OH}$  or ammonia, followed with 200  $\mu\text{l}$  acetonitrile. TiIMAC beads were equilibrated for 2 min with gentle mixing in 200  $\mu\text{l}$  loading buffer consisting of 1 M glycolytic acid 80% acetonitrile and 5% trifluoroacetic acid (TFA). Dried samples were resuspended in 100  $\mu\text{l}$  loading buffer, added to TiIMAC beads, and the mixture was incubated with gentle mixing for 20 min at room temperature (RT). Samples were then washed for 2 min successively with 200  $\mu\text{l}$  loading buffer, three times with 200  $\mu\text{l}$  of 80% acetonitrile-1% TFA, and finally with 200  $\mu\text{l}$  of 10% acetonitrile-0.2% TFA. Phospho-peptides were eluted from beads three times using 80  $\mu\text{l}$  of 1% ammonia, and gentle mixing with pH immediately lowered to 2 using 10% formic acid. Eluted phospho-peptides were pooled, dried in speed vac at RT, and stored at  $-80$  °C before LC-MS analysis.

## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells

**LC-MS/MS Analysis**—Analysis of peptides for total proteome and phospho-proteome was performed on a Velos-Pro orbitrap (Thermo Scientific) mass spectrometer coupled with a Dionex Ultimate 3000 RS (Thermo Scientific). LC buffers were the following: buffer A (2% acetonitrile and 0.1% formic acid in distilled, deionized water (v/v)) and buffer B (80% acetonitrile and 0.08% formic acid in distilled, deionized water (v/v)). All fractions from both total proteome and phospho-proteome were reconstituted in 50  $\mu$ l of 1% formic acid. An aliquot (10  $\mu$ l of total proteome; 15  $\mu$ l of phospho-proteome) of each fraction was loaded at 10  $\mu$ l/min onto a trap column (inner diameter: 100  $\mu$ m  $\times$  length: 2 cm, PepMap nanoViper C18 column, particle size: 5  $\mu$ m, pore size: 100 Å, Thermo Scientific) equilibrated in buffer A for 19 min. The trap column was washed for 6 min at the same flow rate, and then the trap column was switched in-line with a Thermo Scientific, resolving C18 column (inner diameter: 75  $\mu$ m  $\times$  length: 50 cm, PepMap RSLC C18 column, particle size: 2  $\mu$ m, pore size: 100 Å) kept at a constant temperature of 50 °C. Peptides were eluted from the column at a constant flow rate of 300 nL/min with a linear gradient from 5% buffer B to 35% buffer B within 124 min. The column was then washed for 20 min at 98% buffer B and re-equilibrated in 5% buffer B for 19 min. LTQ-Orbitrap Velos Pro was operated in data-dependent positive ionization mode (DDA). The source voltage was set to 2.6 Kv, and the capillary temperature was 250 °C.

A scan cycle comprised MS1 scan (*m/z* range from 335 to 1800) in the velos pro-orbitrap followed by 15 sequential dependent MS2 scans (the threshold value was set at 5000, and the minimum injection time was set at 200 ms) in LTQ with CID. The resolution of the Orbitrap Velos was set at 60,000 after the accumulation of 1,000,000 ions. Precursor ion charge state screening was enabled, with all unassigned charge states and singly charged species rejected. Multistage activation for neutral loss ions was activated only for analysis of phospho-peptides. The lock mass option was enabled for survey scans to improve mass accuracy. To ensure mass accuracy, the mass spectrometer was calibrated on the first day that the runs were performed.

**Peptide and Protein Identification and Quantification**—The MaxQuant setup and parameters for SILAC were consistent with the LFQ experiment described in the previous section with several differences: (i) variable modifications, methionine oxidation, protein N-acetylation, gln  $\rightarrow$  pyro-glu, Phospho (STY); (ii) database: Uniprot-human\_dec2017 (release date of sequence database searched: 12.2017; number of entries: 20,244); (iii) "heavy" label: R10K8, "medium" label: R6K4. Peptide ratios were calculated for each arginine- and/or lysine-containing peptide as the peak area of labeled arginine/lysine divided by the peak area of nonlabeled arginine/lysine for each single-scan mass spectrum. Peptide ratios for all arginine- and lysine-containing peptides sequenced for each protein were averaged. Data were normalized using 1/median ratio value for each identified protein group per labeled sample. Phospho-peptides were normalized using the nonphospho protein 1/median values to correct for mixing errors and compared against the individual nonphospho protein ratio itself to correct for protein regulation interactions. Different parameters used in the iBAQ quantification were: (i) variable modifications—methionine oxidation, protein N-acetylation, Phospho (STY), deamidation (NQ); (ii) database: Homo\_sapiens.GRCh38.pep.all (release date of sequence database searched: 06.2018; number of entries: 107,844); (iii) MS/MS tolerance: FTMS- 20 ppm, ITMS- 0.5 Da.

Valid SILAC quantification was defined as when two out of three replicates were generated with valid SILAC ratios. Valid phospho-proteomic quantification was filtered by localization probability >0.75 in all three replicates.

### *Transient Knockdown With siRNA and Quantitative Reverse Transcription PCR (RT-qPCR)*

Cells were seeded either at 60,000 cells/well in 48-well plates or at 480,000 cells/well in 6-well plates. Prior to cell seeding, plates were coated with respective control siRNA (Silencer Select Negative Control, 4390843), GFPT2 target siRNAs (Silencer Select siGFPT2, s19305 and s19306), GSK3B target siRNAs (Silencer Select siGSK3B, s6239 and s6241), and RELA target siRNAs (Silencer Select siRELA, s11914 and s11915) as well as Lipofectamine RNAiMAX Transfection Reagent (Thermo). Cells were transfected at 37 °C and 5% CO<sub>2</sub> for 48 h with a final siRNA concentration of 10 nM.

In the RT-qPCR experiments, cells were mainly cultured in 48-well plates for 72 h, followed by total RNA extraction with TRI Reagent Solution (Invitrogen). RNA concentration was determined in NanoDrop One (Thermo). In total, 1000 ng of RNA was used for cDNA synthesis on the thermal cycler (MJ research, PTC-225, Peltier Thermal Cycler) using High-Capacity cDNA Reverse Transcription Kit (Thermo). Gene expression was measured with SYBR Green (Luna Universal qPCR Master Mix, NEW ENGLAND BioLabs) on Bio-Rad CFX384 Touch Real-Time PCR Detection System (Bio-Rad). Primers were selected either based on literature or from PrimerBank or designed on Primer3Plus website. Primer sequences for genes in this study (TAG Copenhagen) were listed in [supplemental Table S1](#). The VIM primers were from IDT (Hs.PT.58.38906895).

### *Western Blot*

Cells were incubated with siRNAs as described above. Protein lysates were extracted with RIPA buffer (Pierce, 89900, Thermo) supplemented with protease and phosphatase inhibitors (Halt, 1861284, Thermo) and quantified with BCA protein assay. Proteins were separated on 4 to 12% Bis-Tris gels (NuPAGE, Thermo), transferred to polyvinylidene fluoride (PVDF) membranes (IPFL00010, Immobilon), and probed with antibodies against O-GlcNAcylation (1:200 dilution; sc-59623; Santa Cruz Biotechnology) and the loading control,  $\beta$ -actin (1:2000; MA5-15739; Thermo). The Western blot detection reagents were Clarity Max Western ECL substrate (Bio-Rad), and plots were imaged in the Molecular Imager ChemiDoc XRS+ Systems (Bio-Rad).

### *Proliferation, Scratch, and Invasion Assay*

**Proliferation Assay**—Cells in quadruplicates were seeded at 10,000 cells/well in 96-well plates. GFPT2 knockdown followed the methods described above. For D492 and D492M, 24 h after seeding (48 h for D492HER2), cells were placed under the microscope (LEICA CTR 6500, bright field, 10 $\times$ ) with 5% CO<sub>2</sub> at 37 °C for real-time monitoring and multiple data acquisition. This was controlled by software Micro-Manager 1.4.22. Three spots were chosen in each well, and photos were taken every 6 h. Cell growth was monitored for 66 h for D492 and D492M while 42 h for D492HER2. Photos were batch-processed with Macro in software ImageJ 1.52p, and cell numbers were normalized to the starting time point under the microscope.

**Scratch Assay**—The scratch assay was performed in the IncuCyte ZOOM system (2018A) following the manufacturer's instructions. Cells in triplicates were seeded at 40,000 cells/well in 96-well plates (Essen bioscience, ImageLock, 4379). GFPT2 knockdown followed the methods described above. Cells were scratched and put into the IncuCyte after 48 h of transfection with siRNAs. The IncuCyte ZOOM system took pictures every 2 h. Two positions in each well were chosen, and cells were monitored for 72 h to reach full wound closure. Images were analyzed in the software IncuCyte ZOOM (2018A), and wound confluence data were exported.

**Invasion Assay**—The D492HER2 cells were cultured with siRNA transfection (Scramble and siGFPT2) for 48 h in a 6-well plate. GFPT2 knockdown followed the methods described above. Cells were then reseeded into filter units (Falcon Permeable Support for 24-well Plate with 8.0  $\mu$ m Transparent PET Membrane, 353097) coated with Matrigel (Corning Matrigel Matrix, 356234) at a density of 30,000 cells/well. First, the filter inserts were coated with 100  $\mu$ l 1:10 diluted Matrigel for 20 to 30 min at 37 °C. Next, 300  $\mu$ l of cell suspension was added on top of the filter units. Then, 500  $\mu$ l of H14 medium with 10% FBS was added to the wells in the 24-well plates below the filters. Finally, cells were incubated at 37 °C and 5% CO<sub>2</sub> for 48 h. Noninvasive cells on top of the filters were removed with cotton swabs, followed by fixation with paraformaldehyde (PFA, 3.7%, Sigma, 252549) and DAPI staining (1:5000, Sigma, D9542). Ten images per filter unit were taken by the EVOS FL Auto Imaging System (10 $\times$ , Thermo), followed by the batch analysis of the images in Macro ImageJ 1.52p. To normalize the different cell numbers in the filter units, cells were seeded into a 24-well plate along with the filter units, and they were cultured and treated in the same way as cells in the filter units.

#### Metabolomics Analysis

In the GFPT2 knockdown experiments, cells in triplicates were transfected with control siRNA (scramble), target siRNA (siGFPT2), or neither (wide-type cells) for 48 h in 6-well plates, then cultured for another 24 h before metabolite extraction. In <sup>13</sup>C labeling experiments, wide-type cells were cultured in T25 flasks in triplicates, and after cells reached 80% confluency, the medium was changed to ones without glucose or glutamine. After culturing cells in the medium without glucose or glutamine for 4 h (as Time 0), labeled <sup>13</sup>C 1,2-glucose or <sup>13</sup>C 1(5)-glutamine was added. Metabolites were extracted at time 0 and after 6 h. Metabolites were extracted with cold 80% MeOH supplemented with metabolite internal standards as instructed in an in-house protocol. Extracts were analyzed on the UPLC mass spectrometry (SYNAPT G2, Waters) according to published protocols (29). For <sup>13</sup>C labeling experiments, data were analyzed in ISOCORE, and we normalized the mean enrichment of <sup>13</sup>C in UDP-N-acetylglucosamine (UDP-GlcNAc) to the total amount of UDP-GlcNAc and presented it as relative <sup>13</sup>C incorporation. Metabolomic data were normalized to protein levels.

#### Hydrogen Peroxide (H<sub>2</sub>O<sub>2</sub>) and Reduced Glutathione (GSH) Treatment and Growth Factor Deprivation

The MDA-MB-231 cells were seeded in 24-well plates at 300,000 cells/well and cultured for 48 h followed by treatment with 2  $\mu$ M hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>, Honeywell, 18304H) for 2 h. GFPT2 gene expression was tested by RT-qPCR.

The MDA-MB-231 cells were seeded in 24-well plates at 200,000 cells/well and cultured for 24 h, followed by treatment with 50 mg/l reduced glutathione (GSH, Sigma, G4251) for 48 h. Cells were changed with fresh GSH medium 2 h before the RNA extraction. GFPT2 gene expression was tested by RT-qPCR.

The MDA-MB-231 cells were cultured in the H14 medium as described for the D492 cell lines, then seeded in 24-well plates at 200,000 cells/well and cultured for 24 h followed by treatments with medium deprived of insulin or EGF for 48 h. Fresh medium was changed for the cells 2 h before the RNA extraction. GFPT2 gene expression was tested by RT-qPCR.

#### Glutathione Assay

The glutathione levels, including both reduced (GSH) and oxidized (GSSG) glutathione, were measured with the GSH/GSSG-Glo Assay from Promega (V6611). Cells in quadruplicates were seeded at 20,000 cells/well in 96-well plates. GFPT2 knockdown and H<sub>2</sub>O<sub>2</sub>

treatments followed the methods described above. The glutathione levels were measured 24 h after changed medium. The luminescence signal was detected in the microplate reader (SpectraMax M3, Molecular Devices) with white and opaque 96-well plates (BRANDplates, 781965). To normalize the glutathione level, cells were counted using a crystal violet assay. In short, cells were fixed with 100% cold MeOH and stained with 0.25% crystal violet (Merck, C.I. 42555). After washing, stained cells were dissolved into 100  $\mu$ l of 10% acetic acid and measured at 570 nm in the microplate reader (SpectraMax M3, Molecular Devices LLC).

#### Experimental Design and Statistical Rationale

We conducted two types of proteomics analysis (single-shot LFQ and SILAC with ten fractions) to increase the validity and reproducibility of our results. In the LFQ experiment were four different cell lines (D492, D492M, D492HER2, and D492DEE) in three biological replicates with 12 samples analyzed and described to yield statistical significances. In the SILAC experiment were three different cell lines (D492, D492M, and D492HER2) in three biological replicates with nine samples analyzed and described due to the maximum labeling capacity in SILAC. In both LFQ and SILAC, the epithelial D492 cells were used as controls for the two mesenchymal cell lines. Statistical analysis for all the comparisons between different treatments was conducted in R (two-sided one or two sample(s) Student's *t* test) for the (phospho)proteomic data, metabolomic data, and functional analyses. All error bars represent standard deviation (SD).

The heatmaps and dendrogram were generated in R with packages "ComplexHeatmap," "ggdendro," and "dendextend" (30, 31). Volcano plots were plotted in R with data analyzed in Perseus (version 1.6.2.3, Replace missing values from normal distribution, two-sided Student's *t* test for LFQ, one sample *t* test for SILAC, Permutation-based FDR). GO annotation was performed in Perseus (version 1.6.12.0, Fisher exact test, Benjamini-Hochberg FDR) (32) using all identified proteins from the SILAC experiment as background. We used the R package "pathfindR" (100 iterations; Protein-protein interaction: Biogrid; *p*-values adjustment: "bonferroni," adjusted *p*-value threshold: 0.05) (33) to enrich KEGG pathways. Reactome metabolic pathways were enriched with the default parameters on the Reactome website (Version 65, 67, and 72 were used for D492 versus D492M, D492 versus D492HER2, and D492M versus D492HER2, respectively) (34) and plotted as treemaps in R. The protein interaction networks of proteins involved in the metabolic pathways (enrichment FDR < 0.05) were created in STRING (Version 11.0; *k*-means clustering, minimum required interaction scores: medium confidence 0.400) (35) and visualized in Cytoscape (version 3.5.1/Version 3.6.1) (36). Survival analysis in breast cancer patients was performed in R with packages: "survminer" and "survival". The top and bottom 20th percentile of patients were included in the analysis. Breast cancer patients' data were acquired via the Cancer Genome Atlas (TCGA) cBioPortal (Breast Invasive Carcinoma (TCGA, Provisional)) (37). EMT markers were referenced to and downloaded from the online EMT database (38). RNA expression data of GFPT2 in breast cell lines and breast cancer patients had referred to the Cancer Cell Line Encyclopedia (CCLE) database (39), the Harvard Medical School (HMS) LINC database (40), and TCGA cBioPortal (Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)) (41), respectively. The scatter plots were plotted in R for proteins identified and quantified in both LFQ and SILAC (Pearson). Pathway enrichment of the phosphoproteomic data was performed by Ingenuity Pathway Analysis (IPA) (QIAGEN, version from 2018), while motif enrichment was done in the software Perseus. The R codes for figure plotting can be found on [https://github.com/QiongW56/GFPT2\\_Publication\\_2021](https://github.com/QiongW56/GFPT2_Publication_2021).



RESULTS

*The D492 and D492M Cell Lines Have Basal-like Proteomic Fingerprints, While D492HER2, Closer to D492M, Is Classified as Claudin-low*

The proteomes of the D492, D492M, and D492HER2 cell lines were investigated in biological triplicates by single-shot LFQ and SILAC proteomics with ten fractions (Fig. 1, A and B and supplemental Table S2). In the LFQ experiment (supplemental Data 1), we identified 28,766 peptides corresponding to 3595 protein groups (FDR < 1%). An increase of identified peptides (on average 68,692) was observed in the SILAC experiment due to the fractionation process (supplemental Data 2 and supplemental Fig. S1, A–C). The increased number of peptides in SILAC led to almost twofold more identified (FDR < 1%) and quantified proteins (5120 proteins) compared with that in the LFQ experiment (2705 proteins). The Pearson correlation between LFQ and SILAC was 0.685, 0.782, and 0.847 for ratios of D492M and D492HER2, ratios of D492HER2 and D492, and ratios of D492M and D492, respectively (supplemental Fig. S1, D–F). Cluster analysis showed more similarity between the proteomes of D492M and D492HER2 than to that of D492 (supplemental Fig. S1, G and H and supplemental Data 3). Furthermore, a comparison of unique proteins in D492 versus D492M revealed that the coverage of the proteome was altered by approximately 6.8% to switch between the two cellular phenotypes. For D492 versus D492HER2, this number was approximately 7.0%, and for D492M versus D492HER2, it was about 5.1% (supplemental Table S2).

The proteomic fingerprints of the D492, D492M, and D492HER2 cells were compared with the EMT gene expression signatures reported by Groger *et al.*, 2012 (42) (Fig. 1C). Twenty-six of the 130 reported EMT markers (42) were identified in both the LFQ and SILAC datasets. All but ALDH1A3 were consistent between this study and the literature (supplemental Data 4). D492M clustered with D492HER2, although clear differences were observed in the selected markers between D492M and D492HER2 (Fig. 1C). These proteins represented EMT markers whose expression was inconsistent in the tumorigenic versus nontumorigenic mesenchymal-like cell models. Particularly, genes in clusters C2 and C4 showed different trends in the two mesenchymal cell types. D492HER2 and D492M possessed different mesenchymal characteristics confirmed by the dbEMT2 database (43) (supplemental Fig. S2 and supplemental Data 5). Similar results were obtained when the datasets were compared with the mesenchymal metabolic signatures reported in Shaul *et al.*, 2014 (44) (Fig. 1D and supplemental Data 4).

To position the D492 EMT model in relation to other cell models of breast epithelium, we compared the D492, D492M, and D492HER2 proteomes with the fingerprints of breast cancer reported by Lawrence *et al.*, 2015 (45) (supplemental

Data 6 and Fig. 1E). Both the SILAC and LFQ data placed D492 and D492M with basal-like breast cell lines while D492HER2 clustered with “mesenchymal-like/claudin-low” cell lines.

*Changes to Nucleoside Metabolism Accompany Nontumorigenic and Tumorigenic Mesenchymal Phenotypes in D492*

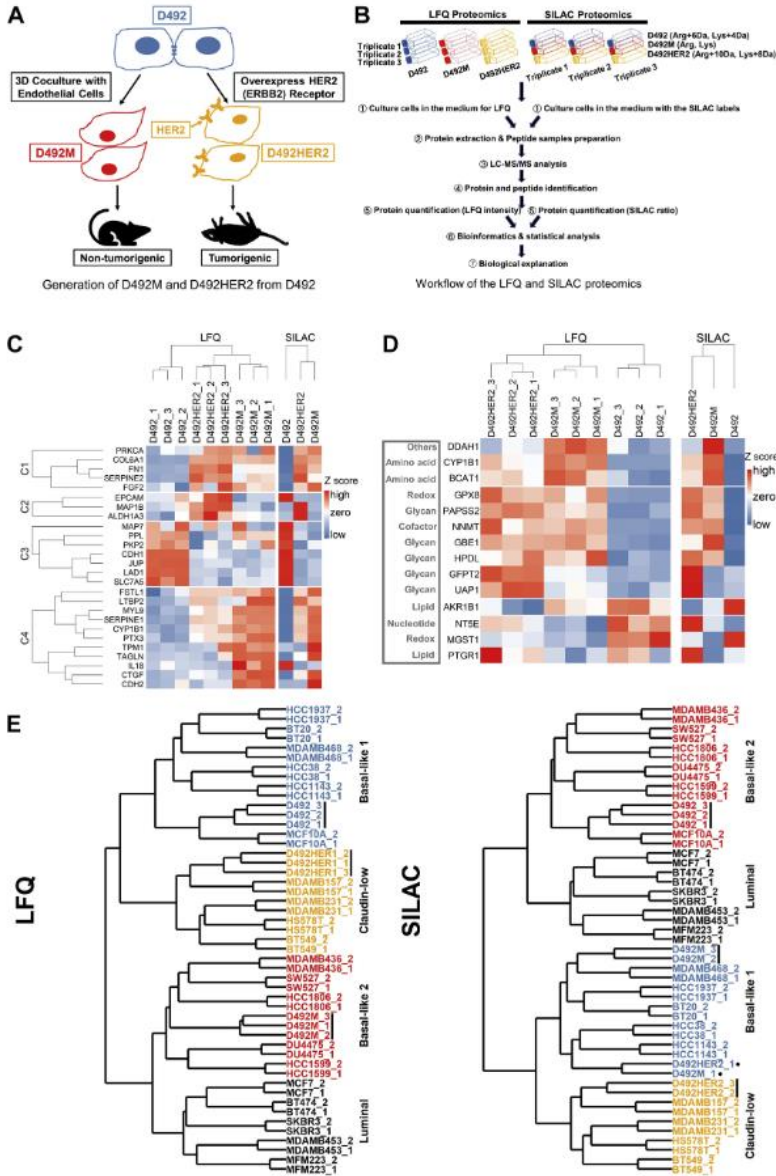
To identify specific proteins different between spontaneous nontumorigenic and the HER2-induced tumorigenic mesenchymal states, protein ratios and *p* values calculated for proteins in both the LFQ (Student's *t* test, two-sample tests, Permutation-based FDR < 0.05) and SILAC (Student's *t* test, one-sample tests, *p* value < 0.05) experiments were plotted for comparison (Fig. 2). Significantly deregulated proteins between the D492 epithelial phenotype and the two mesenchymal phenotypes shared in both the LFQ and SILAC datasets (supplemental Data 7) were analyzed by enrichment analysis of GO terms within Perseus (32) (supplemental Fig. S3, A–C). Of the identified GO terms, 11 were metabolic processes in HER2-induced tumorigenic EMT, while five metabolic processes were enriched in nontumorigenic EMT. Nucleotide-sugar metabolic process was enriched in both comparisons and was also different between the two mesenchymal cell lines. KEGG pathway analysis was also performed to complement these findings. Pathways involved in cell structure, migration, adhesion, invasion, and proteoglycans were enriched in the mesenchymal phenotypes compared with the epithelial phenotype. Nucleoside metabolism altered specifically between the two mesenchymal cell lines (supplemental Fig. S3, D–F and supplemental Data 8).

Focusing on metabolism, we mapped 102 (D492HER2 versus D492M), 84 (D492HER2 versus D492), and 119 (D492 versus D492M) differentially expressed metabolism-related proteins to their respective metabolic pathways (Fig. 3, A–C and supplemental Fig. S4). Asparagine N-linked glycosylation, glycolysis, glucose metabolism, and translocation of SLC2A4 (GLUT4) to the plasma membrane were dysregulated in both mesenchymal transitions. In the HER2-induced mesenchymal model specifically, dysregulation had enriched metabolism in mitochondria. In the nontumorigenic mesenchymal model, different metabolic pathways were enriched, e.g., regulation of ornithine decarboxylase (ODC), selenocysteine synthesis, selenoamino acid metabolism, and metabolism of polyamines. Considering the differences in nucleoside metabolic pathways, the Golgi system, proteoglycans in cancer, and asparagine N-linked glycosylation, we focused our analysis specifically on metabolic proteins involved in glycan metabolism.

*Metabolic Differences in Two Mesenchymal States Involve Changes to GFPT2 Expression*

To determine how changes in proteins involved in glycan precursor synthesis come about following EMT, we identified

## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells



**FIG. 1. Overview of the study and the D492 EMT cell model.** *A*, the process of generating D492M and D492HER2 from D492. The D492 cells were cocultured with endothelial cells (BRENCs or HUVECs) to generate spindle colonies that were subcultured to generate a new cell line D492M. The D492M cells are nontumorigenic. The HER2 (ERBB) receptor was overexpressed on the D492 cells to generate the D492HER2 cell line, and these cells can form tumors in mice. *B*, an overview of the whole proteomic experimental setup in this study from cell



## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells

metabolic enzymes with a fold change of two or more in both the LFQ and SILAC datasets across the three cell lines. These targets were grouped into six clusters that spanned several metabolic pathways and included enzymes that have previously been associated with EMT (Fig. 3, D–F and Table 1). Enzymes closely involved in the metabolism of glycan precursors included PYGB, PGM3, UGDH, PGM2L1, GALNT7, GFPT2, and GALE and were found indiscriminately within different clusters. Altered expression of GALE, UGDH, PGM2L1, and GFPT2 was confirmed using qPCR (Fig. 4A and supplemental Fig. S5, A–C). The biggest difference in RNA expression was detected in GFPT2, consistent with the proteomic analysis (Fig. 4B). GFPT2 was not upregulated in D492 transfected with empty vectors (D492DEE) compared with D492HER2 (Fig. 4C). GFPT2 is the rate-limiting enzyme in the HBP and catalyzes fructose-6-phosphate to glucosamine-6-phosphate while converting glutamine into glutamate. GFPT2 regulates the availability of precursors for O-GlcNAcylation. Knockdown of *GFPT2* with siRNAs reduced protein O-GlcNAcylation in the D492 cells (supplemental Fig. S5, D–F).

### *GFPT2 Influences the EMT Program, Cell Growth, and Cell Invasion, and It Is Associated with Claudin-low Breast Cancer*

The switch between the epithelial marker, E-Cadherin (CDH1), and the mesenchymal marker, N-Cadherin (CDH2), along with the increased expression of vimentin, is hallmark in EMT (46, 47). To interrogate if *GFPT2* knockdown influences the EMT program, we assayed vimentin VIM (Fig. 4, D–F) and the surface markers CDH1 and CDH2 (Fig. 4, G–I and supplemental Fig. S6, A–F). siRNA-mediated knockdown of *GFPT2* decreased the expression of VIM in the mesenchymal cell states and affected the CDH2-to-CDH1 ratios in all three cell lines. Knockdown of *GFPT2* negatively affected cellular growth in both D492M and D492HER2 (Fig. 4, J–L and supplemental Fig. S7A) as well as invasion in D492HER2 (Fig. 4, M and N and supplemental Fig. S7B). Decreasing trends were observed for migration after *GFPT2* knockdown in the three cell lines (Fig. 4O and supplemental Fig. S7C). No changes were observed in cell morphology (supplemental Fig. S6G). To test the generality of these results, we investigated the expression of *GFPT2* across different breast cancer subtypes in both cell lines and patients. *GFPT2* was not

positively associated with HER2-positive but rather claudin-low breast cancer (Fig. 5).

### *Hexosamine Biosynthesis Is Upregulated Post Both Mesenchymal Transitions and Dependent Upon GFPT2*

We next confirmed changes to glycan metabolic precursors in the D492 model. Metabolomics comparison indicated that D492HER2 was more like D492M than D492 (Fig. 6A). UPLC-MS analysis of the glycan precursor metabolites, namely UDP-glucose (UDP-Glc), UDP-glucuronate (UDP-GlcA), N-acetylglucosamine phosphate (GlcNAc-P), and UDP-GlcNAc, showed only significant changes to the GFPT2 product UDP-GlcNAc (Fig. 6B and supplemental Fig. S8, A–C). Relative intracellular concentrations increased by roughly twofold from D492 to D492M and tenfold between D492 and D492HER2. We confirmed altered metabolic activity by monitoring  $^{13}\text{C}$  isotopologue label enrichment in UDP-GlcNAc from cells grown in media containing 1,2- $^{13}\text{C}$  glucose, 1- $^{13}\text{C}$  glutamine, or 5- $^{13}\text{C}$  glutamine. In context with the relative amount of UDP-GlcNAc,  $^{13}\text{C}$  enrichment in UDP-GlcNAc from 1,2- $^{13}\text{C}$  glucose was increased in both D492M and D492HER2. The data indicated an absolute metabolite flux increase into the HBP from glucose that increased via D492 < D492M < D492HER2 (Fig. 6C and supplemental Fig. S8, D–F). Little or no enrichment was observed in UDP-GlcNAc from 1- $^{13}\text{C}$  glutamine. The m + 1 isotopologue in  $^{13}\text{C}$  enrichment from the 5- $^{13}\text{C}$ -glutamine was decreased in D492M compared with D492 and D492HER2 (supplemental Fig. S8F). Knockdown of *GFPT2* resulted in a clear decrease in the intracellular levels of UDP-GlcNAc in both D492M and D492HER2 (Fig. 6D), consistent with reports of its enzymatic function (48).

### *GFPT2 Is a Marker for Cellular Oxidative Stress*

The HBP is altered by changes to cellular redox potential (49). *GFPT2* may influence GSH through glutamine-derived glutamate (50). In addition to changes to UDP-GlcNAc, knockdown of *GFPT2* resulted in decreased intracellular levels of glutamate (Fig. 6E). We similarly noted a decrease in the intracellular cystathionine levels (Fig. 6F). Both glutamate and cystathionine can serve as precursors in GSH *de novo* synthesis. A gene–metabolite correlation analysis of the NCI60 cancer cell line panel indicated a negative correlation between *GFPT2* and GSH (Fig. 7A). No correlation was observed for oxidized glutathione (GSSG), and none was observed for

culture of D492, D492M, and D492HER2 to the bioinformatic and biological analysis of the LFQ and SILAC proteomic datasets. C, dysregulation of EMT markers in independent and published gene expression studies (GES) of EMT, which focused on different cell types and treatment modalities (42). D, dysregulation of EMT metabolic makers in the D492 cell model compared with the literature. There was a consistency between LFQ (left) and SILAC (right) except for NT5E. SILAC was consistent with the literature. HPDL, AKR1B1, and MGST1 were in an opposite trend compared with the literature. The mesenchymal metabolic signature (MMS) in the literature (44) was referred to in this analysis. For detailed descriptions of each EMT marker mentioned in Figure 1, C and D, please refer to the supplemental Data 4. E, classification of the D492 cell model. Using the iBAQ expression of proteins identified in both literature and this study, D492, D492M, and D492HER2 were clustered with other preclassified breast cell lines (45). LFQ (left) classified D492 as “Basal-like 1” (in blue), D492M as “Basal-like 2” (in red), and D492HER2 as “Mesenchymal-like/claudin-low” (in orange), while SILAC (right) classified D492 as “Basal-like 2” (in red), D492M as “Basal-like 1” (in blue), and D492HER2 also as “Mesenchymal-like/claudin-low” (in orange). The LFQ and SILAC raw data were quantified by the iBAQ quantification method in MaxQuant.

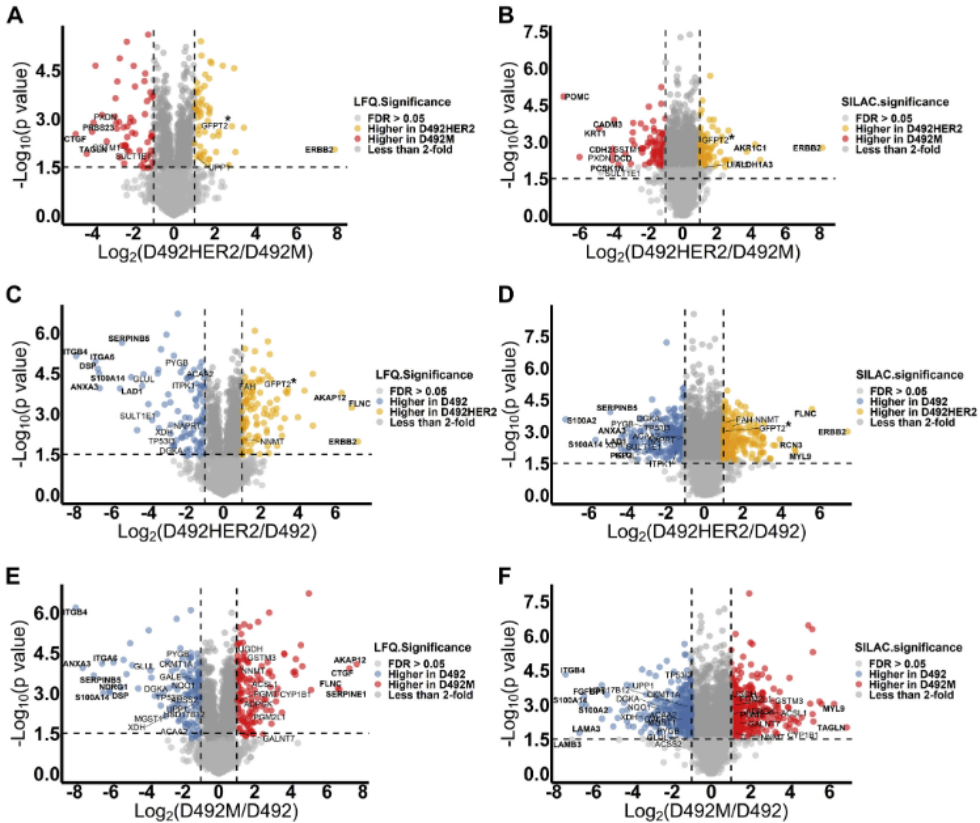


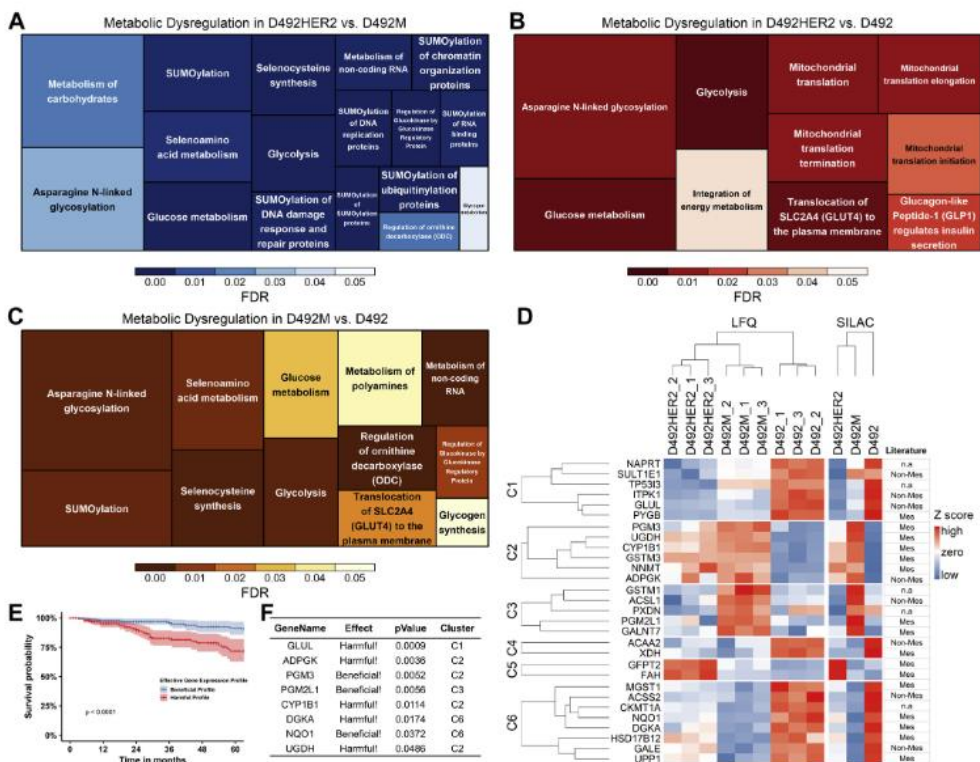
FIG. 2. **LFQ and SILAC proteomic data plotting.** Statistical analysis of the LFQ and SILAC expressions of proteome in two different cell lines: D492HER2 versus D492M (A and B), D492HER2 versus D492 (C and D), and D492M versus D492 (E and F). Proteins with FDR less than 0.05 and fold change of more than 2 were colored. Metabolic enzymes differently expressed in two cell lines and consistent between LFQ and SILAC were labeled in the plots, and proteins with significant differences and big fold changes (at least fourfold between D492HER2 and D492M, fourfold between D492HER2 and D492, and sixfold difference between D492M and D492) were marked in bold. Horizontal dash line indicated  $-\text{Log}_{10}(\text{p value})$  at 1.5, and vertical dash lines indicated fold change at twofold. Proteins with the biggest differences between D492HER2 and D492M were PRSS23, CTGF, TAGLN, POMC, CADM3, KRT1, CDH2, DCD, PCSK1N, AKR1C1, ALDH1A3, and ERBB2, involved in cell adhesion and metabolism. Proteins differently expressed in D492HER2 and D492 were AKAP12, FLNC, ERBB, RCN3, MYL9, SERPINB5, ITGB4, ITGA6, DSP, S100A14, S100A2, LAD1, ANXA3, and PKP2, which were mainly involved in cell adhesion, structure, cell-cell interaction, and signaling. Lastly, a group of proteins that were similar to the differences observed with the other cell lines were differently expressed between D492M and D492, including AKAP12, CTGF, FLNC, SERPINE1, MYL9, TAGLN, ITGB4, ITGA6, ANXA3, SERPINB5, NDRG1, DSP, S100A14, FGFBP1, S100A2, LAMA3, and LAMB3. The main target in this study GFPT2 was highlighted with "\*" in the plots (A–D).

GFPT1 (not shown). Knockdown of *GFPT2*, however, resulted in an increase or no change to glutathione, which we confirmed in the widely studied claudin-low MDA-MB-231 cell line (supplemental Fig. S9, A–H).

Zitzler *et al.* (51) reported that overexpression of GFPT2 enhances cell survival following  $\text{H}_2\text{O}_2$  treatment. We treated MDA-MB-231 cells with  $\text{H}_2\text{O}_2$  and observed an increase of

GFPT2 RNA expression (Fig. 7B) with a concomitant reduction to GSH while total glutathione remained unchanged (Fig. 7, C and D). Furthermore, treatment of the MDA-MB-231 cells with GSH resulted in decreased expression of GFPT2 (Fig. 7E), confirming that GFPT2 expression reacts to GSH. The D492 cells possessed higher amounts of GSH than D492M and D492HER2 (Fig. 7F),

## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells



**FIG. 3. Dysregulated metabolic pathways in two mesenchymal transitions and the metabolic targets identified in this study.** A-C, Reactome metabolic pathways were differently enriched in D492HER2 versus D492M (A), D492HER2 versus D492 (B) and D492M versus D492 (C). Proteins involved in each Reactome metabolic pathway were plotted in [supplemental Fig. S4](#). (Data used for analysis: [Supplemental Data 7](#); Student *t* test, Permutation-based FDR < 0.05, one sample *t* test, *p* value of SILAC ratio < 0.05), differentially expressed metabolic proteins in D492HER2 versus D492M, D492HER2 versus D492, and D492M versus D492 (student *t* test, permutation-based FDR < 0.05 for LFQ, one sample *t* test, *p* value of SILAC < 0.05, more than twofold in both LFQ and SILAC) were manually identified. Samples were in triplicates for both D492, D492M, and D492HER2 in the LFQ and SILAC experiments. For SILAC, the median relative expression of each target for D492, D492M, and D492HER2 was plotted. Targets were clustered into six clusters. The relative expression from the lowest to the highest for each metabolic protein was indicated in color scaling from blue to red, as shown in the color bar. On the right side was listed the identified metabolic targets in mesenchymal (Mes) and nonmesenchymal (Non-Mes) groups based on literature (44). n.a.: not available in literature. E and F, survival analysis of the identified metabolic targets in breast cancer patients of all types revealed a group of enzymes might affect the outcome of patients' survival. The enzymes that exerted either beneficial or harmful effects on patients were listed in the table (F) together with the *p* value for each enzyme's effect on patients. Clusters in which the enzymes resided were listed too. TCGA data, Breast Invasive Carcinoma (TCGA, Provisional), were used in this analysis.

which agreed with the lowest expression of GFPT2 (Fig. 4, A-C).

Cystathionine is an intermediate metabolite in the trans-sulfuration pathway and contributes to hydrogen sulfide (H<sub>2</sub>S) and GSH synthesis. Considering that the levels of cystathionine dropped following GFPT2 knockdown (Fig. 6F) while no consistent significant changes to the glutathione level were observed (supplemental Fig. S9, A-H), we hypothesized

that GFPT2 might affect the intracellular H<sub>2</sub>S homeostasis to counteract oxidative stress. SQOR catalyzes the oxidation of H<sub>2</sub>S and glutathione regenerating ubiquinol in the mitochondrial membrane. Following GFPT2 knockdown, we observed consistent downregulation of SQOR in all four cell lines (Fig. 7, G-N).

NF- $\kappa$ B (p65) responds to cell stress (52) and has previously been shown to modulate GFPT2 (53). Knockdown of p65 did,

TABLE 1  
Metabolic targets identified by comparing one cell line to another

Protein ID	Protein names	Gene name	KEGG classification	Log2(D492HER2/D492M)	Possible transcription factors	Citation related to EMT
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	<i>GFP72</i>	Carbohydrate metabolism	1.558	NF-KB; SIRT6; BMP-2	Shaul <i>et al.</i> , 2014 (44); Simpson <i>et al.</i> , 2012 (50); Szymura <i>et al.</i> , 2019 (53); Taparra <i>et al.</i> , 2019 (96); Zhang <i>et al.</i> , 2018 (55); Zhou <i>et al.</i> , 2019 (22)
Q16831	Uridine phosphorylase 1	<i>UPP1</i>	Pyrimidine metabolism	1.167	NF-Kb; Oct3/4	Guan <i>et al.</i> , 2019 (97); Wehbe <i>et al.</i> , 2012 (98)
X5DR03	Glutathione S-transferase mu 1 isoform B (Fragment)	<i>GSTM1</i>	Glutathione metabolism	-2.152	Nrf2	n.a
Q53X91	Sulfotransferase (Fragment)	<i>SULT1E1</i>	Steroid hormone biosynthesis	-2.383	Nrf2	n.a
Q92626	Peroxidase homolog	<i>PXDN</i>	Oxidoreductases	-3.825	Snail 1; Nrf2	Briem <i>et al.</i> , 2019 (16); Sitole and Mavri-Damelin, 2018 (99)
Log2(D492HER2/D492)						
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	<i>GFP72</i>	Carbohydrate metabolism	1.827	NF-KB; SIRT6; BMP-2	Shaul <i>et al.</i> , 2014 (44); Simpson <i>et al.</i> , 2012 (50); Szymura <i>et al.</i> , 2019 (53); Taparra <i>et al.</i> , 2019 (96); Zhang <i>et al.</i> , 2018 (55); Zhou <i>et al.</i> , 2019 (22)
Q6FH49	NNMT protein	<i>NNMT</i>	Nicotinate and nicotinamide metabolism	1.275	Stat3	Eckert <i>et al.</i> , 2019 (100); Shaul <i>et al.</i> , 2014 (44)
P16930	Fumarylacetoacetase	<i>FAH</i>	Tyrosine metabolism	1.272	CDC5L	n.a
Q6XQN6	Nicotinate phosphoribosyltransferase	<i>NAPRT</i>	Nicotinate and nicotinamide metabolism	-1.266	NF-Kb; STAT3; HIF-1a	Lee <i>et al.</i> , 2018 (101)
A0A024R6H3	Inositol 1,3,4-triphosphate 5/6 kinase, isoform CRA_a	<i>ITPK1</i>	Inositol phosphate metabolism	-1.358	BMP2; TBX2; SNAIL; miR-23b	Bonet <i>et al.</i> , 2015 (102)
A0A0B4J2A4	3-ketoacyl-CoA thiolase, mitochondrial	<i>ACAA2</i>	Lipid metabolism	-1.737	PPAR $\alpha$ ; HNF4 $\alpha$	n.a
A0A024RB23	Diaclyglycerol kinase	<i>DGKA</i>	Lipid metabolism	-1.826	PPAR $\gamma$ ; Stat5; AP2, Ets1, SP1	n.a
P11216	Glycogen phosphorylase, brain form	<i>PYGB</i>	Starch and sucrose metabolism	-2.018	n.a	Zhang <i>et al.</i> , 2018 (55)
P47989	Xanthine dehydrogenase/oxidase	<i>XDH</i>	Purine metabolism	-2.908	NF-Y	n.a
Q53FA7	Quinone oxidoreductase PIG3	<i>TP53I3</i>	Oxidative stresses and irradiation	-2.911	FOXK2&BAP1	Alonso <i>et al.</i> , 2007 (103); Reka <i>et al.</i> , 2014 (104)
Q53X91	Sulfotransferase (Fragment)	<i>SULT1E1</i>	Steroid hormone biosynthesis	-3.010	Nrf2	n.a
A8YXX4	Glutamine synthetase	<i>GLUL</i>	Carbohydrate metabolism	-3.582	ATF4	n.a
Log2(D492M/D492)						
Q53TK1	Cytochrome P450, family 1, subfamily B, polypeptide 1, isoform CRA_a	<i>CYP1B1</i>	Lipid metabolism	3.733	SP1	Kwon <i>et al.</i> , 2016 (105); Shaul <i>et al.</i> , 2014 (44)



TABLE 1—Continued

Protein ID	Protein names	Gene name	KEGG classification	Log2(D492HER2/D492M)	Possible transcription factors	Citation related to EMT
E7EPM6	Long-chain-fatty-acid-CoA ligase 1	<i>ACSL1</i>	Lipid metabolism	1.712	SP1	Sánchez-Martínez <i>et al.</i> , 2015 (106)
Q6PCE3	Glucose 1,6-bisphosphate synthase	<i>PGM2L1</i>	Carbohydrate metabolism	1.422	ZEB1	n.a
Q95394	Phosphoacetylglucosamine mutase	<i>PGM3</i>	Carbohydrate metabolism	1.393	n.a	n.a
Q6FGJ9	Glutathione S-transferase	<i>GSTM3</i>	Glutathione metabolism	1.368	Nrf2	Zhou <i>et al.</i> , 2008 (107)
Q86SF2	N-acetylgalactosaminyltransferase 7	<i>GALNT7</i>	Glycan biosynthesis and metabolism	1.328	miR-30d/30b; miR-214	n.a
Q6FH49	NNMT protein	<i>NNMT</i>	Nicotinate and nicotinamide metabolism	1.181	Stat3	Eckert <i>et al.</i> , 2019 (100); Shaul <i>et al.</i> , 2014 (44)
Q9BRR6-2	Isoform 2 of ADP-dependent glucokinase	<i>ADPGK</i>	Glycolysis/ Gluconeogenesis	1.173	n.a	Lee <i>et al.</i> , 2016 (108); Song <i>et al.</i> , 2018 (109)
O60701	UDP-glucose 6-dehydrogenase	<i>UGDH</i>	Carbohydrate metabolism	1.025	SP1	Tang <i>et al.</i> , 2016 (110); Vergara <i>et al.</i> , 2015 (111)
H0UIA1	Acyl-CoA synthetase short-chain family member 2, isoform CRA_c	<i>ACSS2</i>	Carbohydrate metabolism	-1.125	SREBF1/2; HIF; TFEB	Sun <i>et al.</i> , 2017 (112)
Q53GQ0	Very-long-chain 3-oxoacyl-CoA reductase	<i>HSD17B12</i>	Lipid metabolism	-1.142	n.a	n.a
Q53FA7	Quinone oxidoreductase PIG3	<i>TP53I3</i>	Oxidative stresses and irradiation	-1.159	FOXK2&BAP1	Alonso <i>et al.</i> , 2007 (103); Reka <i>et al.</i> , 2014 (104)
Q14376	UDP-glucose 4-epimerase	<i>GALE</i>	Galactose metabolism	-1.458	n.a	n.a
Q6LET6	MGST1 protein (Fragment)	<i>MGST1</i>	Glutathione metabolism	-1.522	n.a	Fischer <i>et al.</i> , 2015 (113); Shaul <i>et al.</i> , 2014 (44)
P11216	Glycogen phosphorylase, brain form	<i>PYGB</i>	Starch and sucrose metabolism	-1.611	n.a	Zhang <i>et al.</i> , 2018 (55)
P12532	Creatine kinase U-type, mitochondrial	<i>CKMT1A</i>	Arginine and proline metabolism	-1.686	LncRNA n335586&miR-924; EVI1&RUNX1	Tanaka and Ogishima, 2015 (114)
A0A0B4J2A4	3-ketoacyl-CoA thiolase, mitochondrial	<i>ACAA2</i>	Lipid metabolism	-2.038	PPAR $\alpha$ ; HNF4 $\alpha$	n.a
B4DLR8	NAD(P)H dehydrogenase [quinone] 1	<i>NQO1</i>	Ubiquinone and other terpenoid-quinone biosynthesis	-2.069	Nrf2; NF-Kb	Fischer <i>et al.</i> , 2015 (113); Yang <i>et al.</i> , 2017 (115)
Q16831	Uridine phosphorylase 1	<i>UPP1</i>	Pyrimidine metabolism	-2.303	NF-Kb; Oct3/4	Guan <i>et al.</i> , 2019 (97); Wehbe <i>et al.</i> , 2012 (98)
A0A024RB23	Diacylglycerol kinase	<i>DGKA</i>	Lipid metabolism	-2.730	PPAR $\gamma$ ; Stat5; AP2, Ets1, SP1	n.a
A8YXX4	Glutamine synthetase	<i>GLUL</i>	Carbohydrate metabolism	-2.741	ATF4	n.a
P47989	Xanthine dehydrogenase/oxidase	<i>XDH</i>	Purine metabolism	-3.159	NF-Y	n.a

These targets were with significance (Permutation-based FDR less than 0.05). They were at least twofold changes, comparing D492HER2 to D492M, D492HER2 to D492, and D492M to D492. The fold changes were confirmed by both LFQ and SILAC. The average of Log2 ratios from LFQ and SILAC were reported in this table.

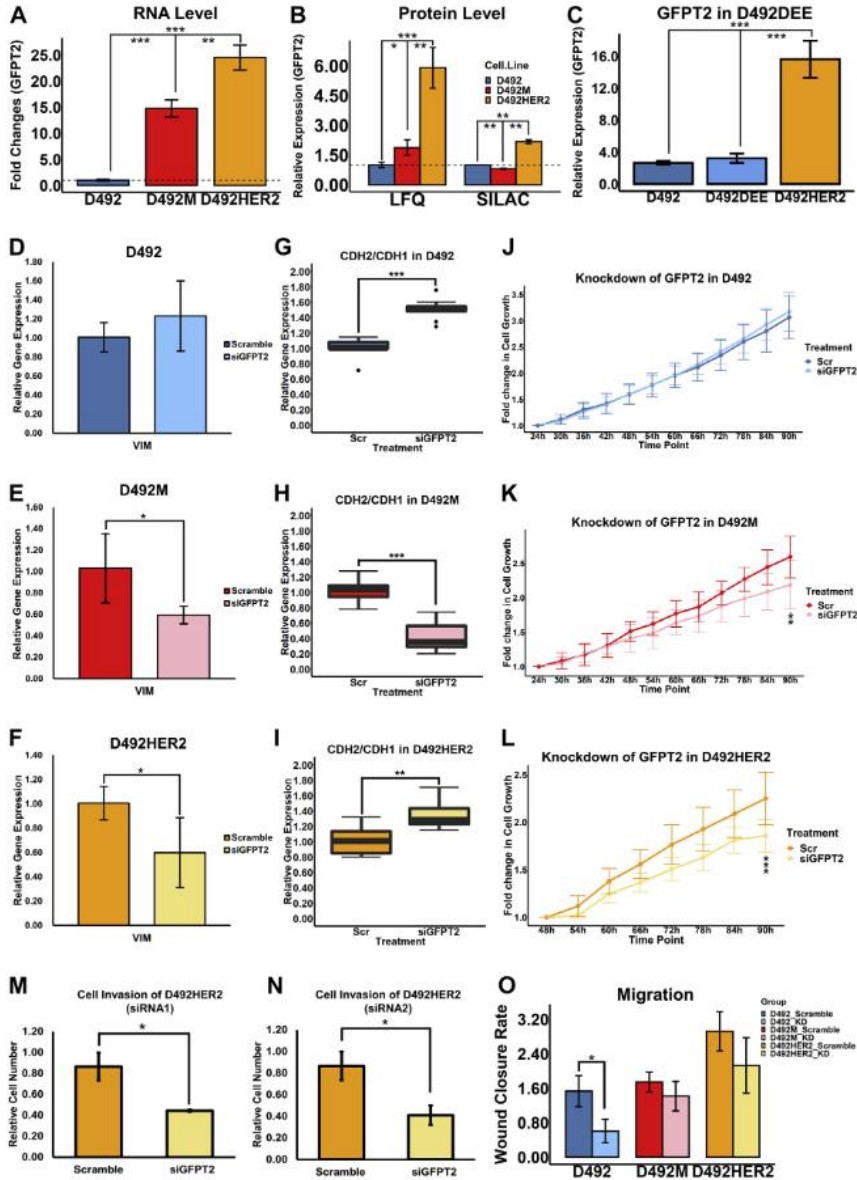


FIG. 4. Expression and functions of GFPT2 in the D492 EMT model. A, GFPT2 showed the highest expression in D492HER2 while lowest in D492 on the RNA level; B, The protein expression of GFPT2 in the three cell lines suggested highest expression of GFPT2 in D492HER2 confirmed by both LFQ (left) and SILAC (right); C, The GFPT2 level in D492DEE which was the negative control cell line of D492HER2 indicated that the increased expression of GFPT2 was not due to the artifacts from cell handling but the overexpression of HER2. D-F, siRNA-mediated

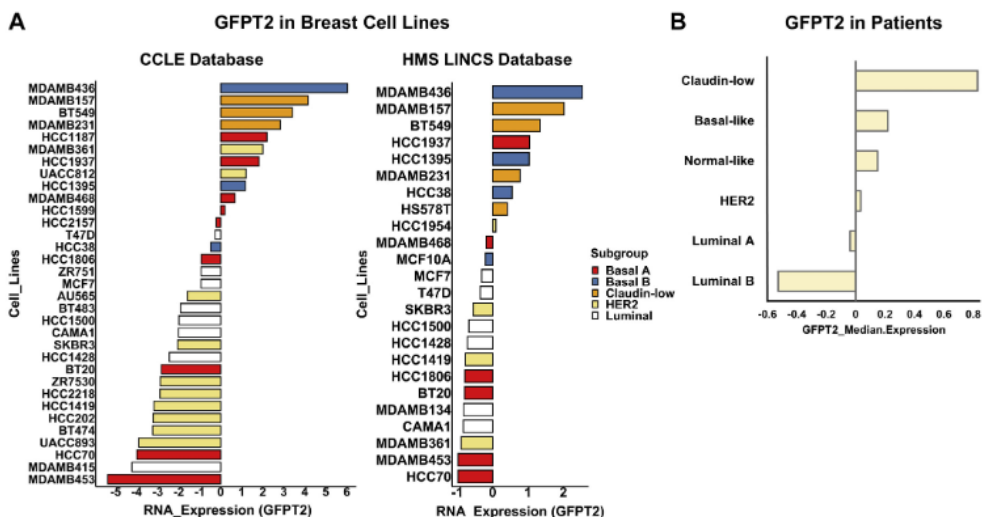


FIG. 5. GFPT2 is higher in basal and claudin-low breast cell lines, and same trend is shown in breast cancer patients. A, based on data from an open-source database – CCLC (left) (39), GFPT2 was higher in basal and claudin-low cell lines while lowly expressed in HER2-positive and luminal cell lines. The same trend was seen with data from another open-source database – HMS LINCS (right) (40). The molecular classification of breast cancer cell lines was based on literature (95). B, TCGA data (Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)) suggested GFPT2 was expressed highly in claudin-low patients, while its expression was lower in HER2-positive and luminal patients.

however, not suppress GFPT2 expression in D492HER2 (supplemental Fig. S9, I and J). Analysis of previously published secretome data from D492 and D492HER2 (54) showed differences in proteins involved in TGF- $\beta$ , IGF, TNF, and EGF signaling (Fig. 8A), with all apart from IGF confirmed to regulate GFPT2 (53, 55, 56). Individual removal of growth factors from MDA-MB-231 growth media resulted in decreased GFPT2 expression following removal of insulin and EGF (Fig. 8, B and C) consistent with receptor tyrosine kinase (RTK) regulation of GFPT2. Expression of the membrane receptor, IGF1R, was also higher in D492HER2 than in D492 (Fig. 8D), supporting the higher activity of IGF signaling in D492HER2. ERK/MAPK are common downstream regulators in the RTK signaling pathways. Phosphoproteomics analysis (supplemental Data 9) confirmed changes in signaling within the ERK/MAPK pathway between D492HER2 and D492 and showed enrichment of the GSK3- $\beta$  and PKC $\alpha$  substrates (Fig. 8, E and F). siRNA-mediated knockdown of GSK3- $\beta$  resulted in increased GFPT2 expression (Fig. 8, G–M).

DISCUSSION

To define changes to metabolic enzymes associated with EMT phenotypes in the breast gland, we analyzed proteins isolated from three breast cell lines representing three epithelial–mesenchymal states using both LFQ- and SILAC-based proteomics mass spectrometry. We first analyzed the proteomics data to confirm the EMT signature of the D492 EMT model and position the cell lines with respect to other cells derived from breast tissue.

The expression pattern of EMT markers was consistent with previously reported markers of EMT (42, 43). Groger *et al.*, 2012 compared EMT gene expression signatures during cancer progression from 18 independent and published papers and listed the core genes involved in EMT. Good consistency between literature and our datasets was observed in terms of the up/downregulation of these EMT markers and between the two detection methods. LFQ and SILAC were discordant on the expression of IL18 and EPCAM. Results from SILAC were more in agreement with literature reports.

knockdown of GFPT2 decreased VIM in both mesenchymal cell lines. G–I, knockdown of GFPT2 affected CDH2-to-CDH1 ratios in all cell lines. J–L, knockdown of GFPT2 decreased the growth of D492M (K) and D492HER2 (L) after 90 h from cell seeding. M and N, D492HER2 cell invasion was decreased after knockdown of GFPT2 and confirmed by two siRNAs. O, D492 cell migration was slowed by the knockdown of GFPT2. A decreasing trend was seen in D492M and D492HER2 without significance. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

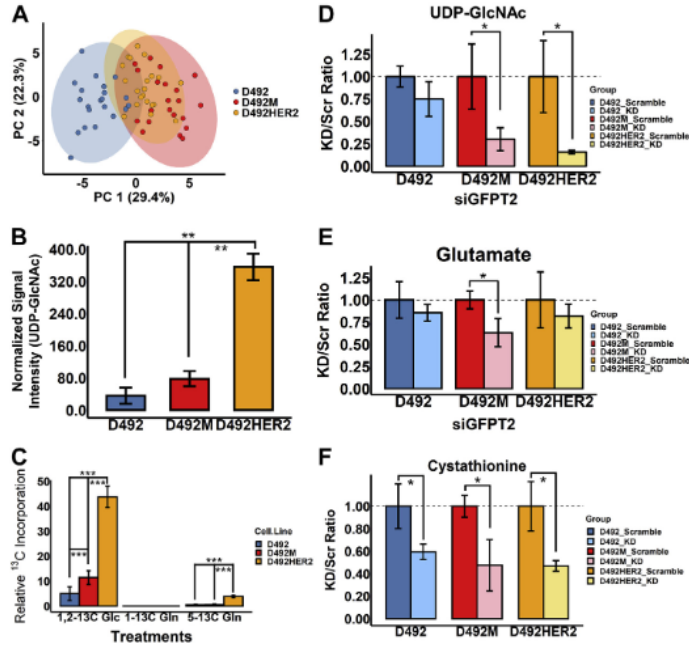


FIG. 6. **Metabolomic analyses of the D492 EMT cell model.** A, the metabolome in D492M and D492HER2 was similar compared with that in D492. B, UDP-GlcNAc was expressed higher in D492HER2 compared to the other cell types. C, carbon incorporation into UDP-GlcNAc after 6 h' culture. Carbons from 1,2-<sup>13</sup>C glucose (Glc) were highly incorporated into UDP-GlcNAc, compared to 5-<sup>13</sup>C glutamine (Gln) and 1-<sup>13</sup>C Gln in all three cell lines. No carbon incorporation from 1-<sup>13</sup>C Gln in all three cell lines. Higher rates of carbon incorporation into UDP-GlcNAc from both 1,2-<sup>13</sup>C Glc and 5-<sup>13</sup>C Gln were observed in D492HER2, compared to D492 and D492M. A higher rate of carbon incorporation into UDP-GlcNAc from 1,2-<sup>13</sup>C Glc was observed in D492M compared with D492. D, knockdown of *GFPT2* decreased the production of UDP-GlcNAc in both D492M and D492HER2. E, a decreasing trend for glutamate with *GFPT2* knockdown was observed in the D492 EMT model. F, knockdown of *GFPT2* significantly decreased cystathionine in all three cell types. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001.

The results support the epithelial and mesenchymal phenotypes of the D492 EMT cell model previously reported (14, 15).

The results define the D492 model better in relation to other cell models used to study breast cancer and EMT. Both D492 and D492M clustered within the “basal-like” categories, consistent with the prior classification of D492 (15). D492 clustered with the human breast epithelial cell line MCF10A that, like D492, is derived from a reduction mammaplasty from patients without breast cancer (57). Both are nontumorigenic, and MCF10A, like D492, expresses stem cell-like markers (58–60). Based on SILAC, D492M was most similar to the tumorigenic cell line MDA-MB-468 (59, 61) originally isolated from a metastatic adenocarcinoma and has been used to study metastasis previously (62). D492HER2 shared more similarities with D492M than D492 based on the proteome clustering but was characterized as claudin-low. Accordingly, D492HER2 thus appears to be an intermediate between D492

and D492M, representing diversion from the natural EMT program upon which tumorigenic properties are gained. Both the LFQ and SILAC data indicated D492HER2 as a “mesenchymal-like/claudin-low” cell type showing the most similarity to the tumorigenic MDA-MB-157 cells originally isolated from metastatic human breast carcinoma (59, 61, 63). Given the basal origin of D492, the relatively small changes to the coverage of the proteome between these cell lines (5–7%), and that the claudin-low phenotype has recently been redefined as a molecular signature found dispersed within the intrinsic breast cancer subtypes (64), these results define D492HER2 as a basal-like cell line with claudin-low phenotypes. The findings position the D492 cell culture model with respect to other commonly researched cell culture models originating from breast tissue based upon their protein content and suggest that the D492 cell model mimics basal-like tumors with D492HER2 prone to claudin-low.



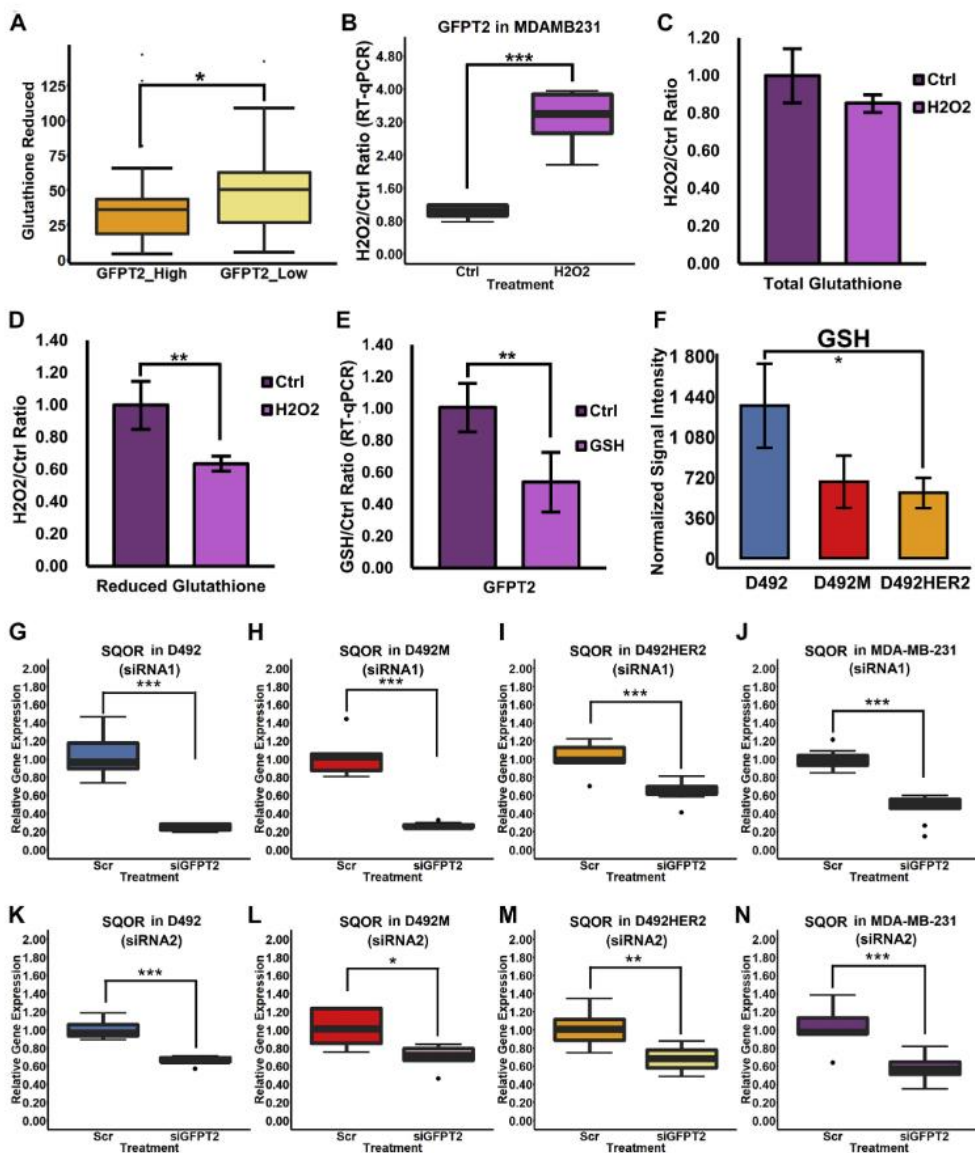


FIG. 7. **GFPT2 is a marker for cellular stress.** A, gene-metabolite correlation analysis of the NCI60 cancer cell line panel indicated a negative correlation between GFPT2 and GSH. B, GFPT2 RNA expression was significantly upregulated with 2  $\mu\text{M}$   $\text{H}_2\text{O}_2$  treatment in MDA-MB-231. C, the total glutathione level did not change after  $\text{H}_2\text{O}_2$  treatment. D, the GSH level was significantly decreased by the  $\text{H}_2\text{O}_2$  treatment. E, treatment with 50 mg/l of GSH significantly downregulated the GFPT2 gene expression in MDA-MB-231. F, GSH level was significantly higher in D492 than in D492M and D492HER2. G-N, SQOR RNA expression was significantly downregulated in D492 (G), D492M (H), D492HER2 (I), and MDA-MB-231 (J) by siRNA-mediated knockdown of GFPT2, which was confirmed by the second siRNA (K-N). \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

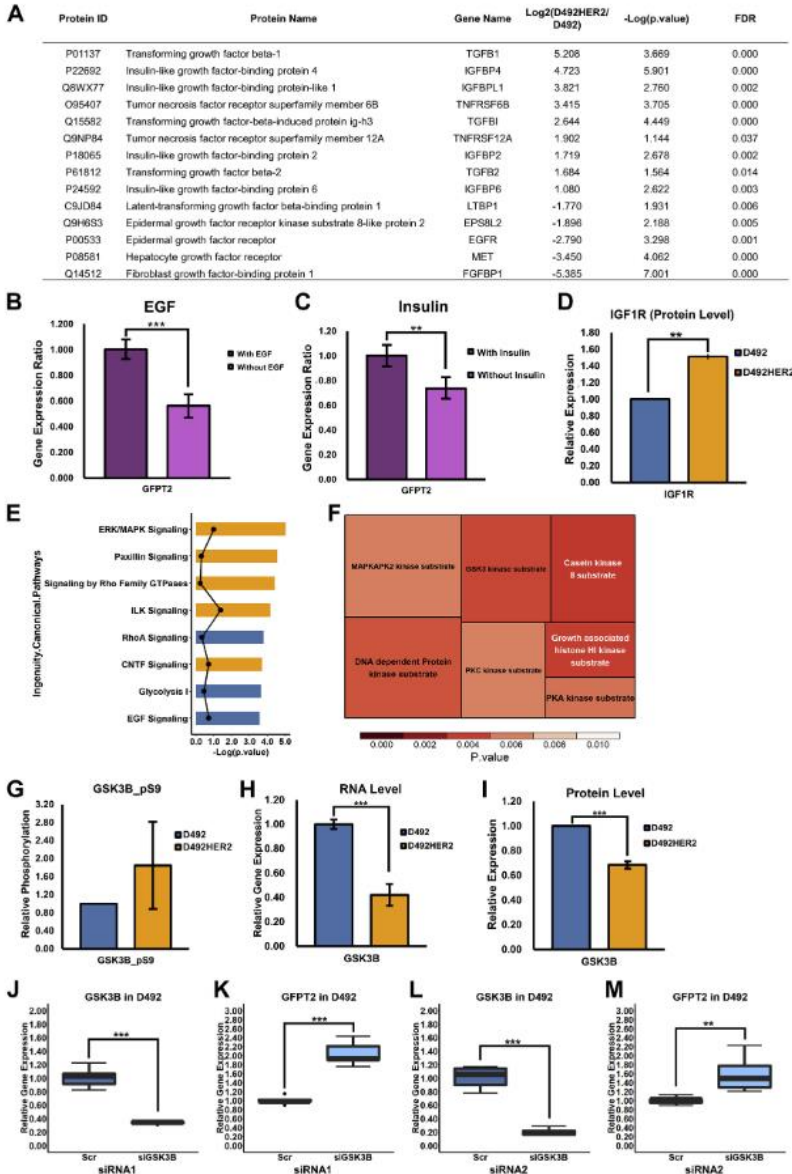


Fig. 8. Signaling regulation of GFPT2. A, secretome of D492HER2 and D492 revealed a list of growth factors that secreted differently between these two cell lines (FDR <0.05, Fold change  $\geq 2$ ). B and C, to test the effects of growth factors on GFPT2, we adapted the MDA-MB-231 cells with the FBS-free H14 medium. Removal of EGF and insulin decreased GFPT2 RNA expression in the MDA-MB-231 cell line. D, the protein level of IGF1R was higher in D492HER2 than in D492 based on the SILAC proteomic data. E, top eight of the Ingenuity Canonical

## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells

Focusing on metabolism, a variety of metabolic enzymes involved in a diversity of metabolic pathways were significantly altered in EMT according to our data, supporting that EMT is entangled with the metabolic network, e.g., central carbon metabolism including glycolysis and oxidative phosphorylation, pentose phosphate pathway (PPP), and mitochondrial metabolism, lipid metabolism, glutamine metabolism, nucleotide metabolism, and glycan metabolism (65). The upregulation of PGM3, UAP1, and OGT, also components of the HBP in the mesenchymal cells, supports the increased activities of HBP in EMT. GO enrichment analysis and pathway enrichment analysis further indicated differences in glycan metabolism in the D492 EMT model. Multiple transcriptional factors, regulators, and enzymes are influenced by O-GlcNAcylation, and glycans are essential for the formation and function of the extracellular matrix (66, 67). GlcNAcylation plays an essential role in breast cancer metastasis and tumorigenesis (68), in line with the observations that siRNA-mediated knockdown of *GFPT2* imparted negative effects on growth and invasion in the mesenchymal cell lines.

GFPT2 has previously been identified as part of the mesenchymal metabolic signature genes (44) and associated with invasive breast cancer mesenchymal phenotypes on the mRNA level (50). Several studies have focused on the function and regulation of GFPT2 related to its role in modulating O-GlcNAcylation of proteins on account of GFPT2 in producing UDP-GlcNAc (21, 22, 53). GFPT2 has also been shown to counteract oxidative stress (51, 69, 70), although the mechanism behind that remains elusive. Our results demonstrate that GFPT2 affects protein O-GlcNAcylation, regulates the EMT program, and impacts cellular growth and invasion in a cellular subtype-specific manner in breast epithelial cells, which are consistent with the literature mentioned above. Claudin-low breast cancer has recently been redefined and subclassified as a breast cancer subtype (64, 71). GFPT2 was one of the predicted claudin-low signatures in Triple-Negative Breast Cancer reported by Prat *et al.*, 2010 (72). KRAS and LKB1 comutant NSCLC emulates claudin-low breast cancer, and GFPT2 was reported in different studies to be the key player in boosting the malignancy of this type of malignant lung cancer (73, 74). Our results indicate that GFPT2 is a claudin-low breast cancer marker, consistent with the previous finding that D492HER2 with higher expression of GFPT2 belongs to the claudin-low breast cell line. The upregulation of GFPT2 in D492HER2 compared with its negative control cell

line D492DEE indicates that the HER2 receptor is somewhat responsible for the GFPT2 overexpression. The lower levels of GFPT2 across HER2-positive cell lines in the public domain however suggest that the HER2 receptor is not the only regulator of GFPT2.

We confirmed increased HBP flux associated with GFPT2 expression. The HBP is central to metabolic rewiring in cancer as it affects glutamine, acetyl-CoA, the nucleotides UTP and UDP, and the glycan substrate UDP-GlcNAc (66). UDP-GlcNAc intracellular concentration increased in accordance with the expression levels of GFPT2 in the D492 cell lines and dropped following *GFPT2* knockdown. Concordantly, <sup>13</sup>C flux analysis showed increased flux from glucose and glutamine into UDP-GlcNAc. The altered glutamine flux profiles into UDP-GlcNAc are consistent with previously proposed differences in TCA cycle flux in the D492 model on account of altered glutamine utilization following EMT (9). The increased metabolic flux observed alongside enhanced expression of GFPT2 is consistent with a mass action effect and corresponds to GFPT2's role as a biomarker for glucose uptake independent of GLUT1 (55, 70). <sup>13</sup>C enrichment from the 5-<sup>13</sup>C-glutamine was negligible but suggestive of flux rerouting in the TCA cycle, particularly in D492M compared with D492 and D492HER2. Specifically, the changes in the m + 1 isotopologue were indicative of alternate carbon contribution to UDP-GlcNAc through citrate-derived cytosolic acetyl-CoA and aspartate and are consistent with more detailed metabolic flux analysis of these cell lines reported in Karvelsson *et al.*, 2021 (23).

In light of increased glutamine uptake following EMT in the D492 cells (9) along with decreased glutamate and cystathionine following *GFPT2* knockdown, we explored if GFPT2 would influence GSH through GFPT2 derived glutamate. Knockdown of *GFPT2* resulted in no change or trends toward increased glutathione that does not support a positive relationship between GFPT2 and glutathione in the four cell lines tested. The regulatory role of GFPT2 on glutathione can however not be excluded merely based on the little impacts of GFPT2 on the net glutathione levels. A negative correlation between GSH and GFPT2 expression was observed across the NCI60 cancer cell line panel. Therefore, even though GFPT2 had limited effects on glutathione, the expression of GFPT2 may be adjusted according to the GSH level. Indeed, H<sub>2</sub>O<sub>2</sub> treatment increased GFPT2 expression, while GSH treatment had the opposite effect. The D492 cells possessed

Pathways from the phosphoproteomics data analysis. Pathways activated in D492HER2 were in orange, while pathways activated in D492 were in blue. Dots referred to the absolute value of activation Z-scores. Pathways were listed based on *p* value. *F*, motif enrichment from Perseus (Version 1.6.14.0) suggested a list of kinases behaving differently in D492HER2 compared to D492. *G*, based on the phosphoproteomics analysis, GSK3- $\beta$  was highly phosphorylated at position serine 9, which inhibits GSK3- $\beta$  activation in D492HER2 compared with D492. *H* and *I*, RNA (*H*) and protein (*I*) expression of GSK3- $\beta$  in D492HER2 versus D492 indicated the higher abundance of GSK3- $\beta$  in D492. *J–M*, knockdown of GSK3- $\beta$  in D492 increased GFPT2 RNA expression. *J*, knockdown efficiency for GSK3- $\beta$  with the first siRNA. *K*, GFPT2 RNA expression after knockdown of GSK3- $\beta$  in D492 with the first siRNA. *L*, knockdown efficiency for GSK3- $\beta$  with the second siRNA. *M*, GFPT2 RNA expression after knockdown of GSK3- $\beta$  in D492 with the second siRNA. \*\**p* < 0.01; \*\*\**p* < 0.001.

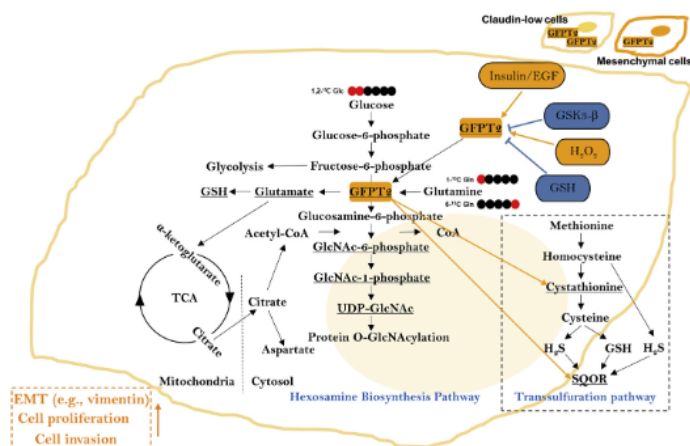


Fig. 9. **Summary of the study.** The hexosamine biosynthesis pathway (HBP) and transsulfuration pathway, two main pathways associated with GFPT2 in this study, were illustrated. HBP is closely associated with the central carbon metabolism. The underscored metabolites were measured in this study. We used 1,2-<sup>13</sup>C Glc, 1-<sup>13</sup>C Gln, and 5-<sup>13</sup>C Gln in the <sup>13</sup>C tracing experiment to trace the carbon flux into UDP-GlcNAc. GFPT2 is the rate-limiting enzyme in the HBP and was upregulated in mesenchymal cells, especially in partial EMT. It affected the HBP flux, EMT program (e.g., VIM), cell growth, and cell invasion, and it was overexpressed in claudin-low breast cancer represented by the D492HER2 cell line. GFPT2 was regulated by oxidative stress (H<sub>2</sub>O<sub>2</sub> and GSH) and signaling regulators (insulin and EGF, and GSK3-β). Knockdown of *GFPT2* with siRNAs decreased the cystathionine level and SQOR RNA expression in the transsulfuration pathway.

higher amounts of GSH while GFPT2 expression was low, while the opposite was true for D492M and D492HER2, consistent with increased cell stress in the mesenchymal cells as previously reported (9, 75). The data suggest that redox balance influences GFPT2 expression. High expression of GFPT2 is a marker for oxidative stress important for EMT (76) and breast cancer progression (77).

NF-κB is central to the cellular stress response and is implicated in EMT (52, 78). Following TGF-β/TNFα stimulation, GFPT2 expression is enhanced by the stress regulator NF-κB with which it forms a regulatory feedback loop via glycosylation of p65 (21). siRNA-mediated knockdown of p65 did, however, not influence GFPT2 expression in the D492 EMT model. Our results do not exclude that NF-κB induces GFPT2 expression in the stress response with TNFα stimulation. However, it appears that maintenance of GFPT2 expression in claudin-low breast cancer relies on additional factors.

H<sub>2</sub>S originates from the transsulfuration pathway, and knockdown of *GFPT2* resulted in decreases to the pathway intermediate, cystathionine, suggesting that the production of H<sub>2</sub>S could be hampered. SQOR utilizes H<sub>2</sub>S as substrate, and decreased SQOR with *GFPT2* knockdown supports the hampered H<sub>2</sub>S production. However, a solid relationship between GFPT2 and H<sub>2</sub>S could not be established based on limited evidence. Nevertheless, the negative effects of knocking down *GFPT2* on cystathionine and SQOR have connected GFPT2 with H<sub>2</sub>S and further with mitochondrial

functions. H<sub>2</sub>S signaling has recently been reported to facilitate EMT (79, 80). However, the exact role of H<sub>2</sub>S in EMT is still controversial (81). Mitochondrial dysfunction is involved both in EMT (82) and in breast cancer (83). A recent study has connected GFPT2 to SLP-2 involved in mitochondrial regulation (70). Understanding the roles of GFPT2 in oxidative stress and H<sub>2</sub>S and mitochondrial homeostasis in mesenchymal cells is beneficial to clinical therapeutic interventions and prognostics. Our results indicate that the effects of GFPT2 on GSH are more complex than can be accounted for by a direct impact on account of limitation to GSH precursors. GFPT2 expression responds to changes in the intracellular redox environment and may alter H<sub>2</sub>S level that impacts SQOR and mitochondrial homeostasis. The protective effects of GFPT2 from oxidative stress may thus be attributed in part to changes in H<sub>2</sub>S-SQOR activity, although further research is needed to elucidate this link.

The regulation of GFPT2 is inherently complex and associated with various growth factors and transcriptional regulators, including EGF, TGF-β, TNF, NF-κB, SIRT6, sXBP1, and SLP-2, etc. (53, 55, 56, 70, 84). Mutant KRAS has also been demonstrated to enhance flux into the HBP via GFPT2 that is potentiated by loss of LKB1 (74). Our results are consistent with growth factor-mediated regulation of GFPT2 as removal of insulin and EGF from growth media decreased the expression of GFPT2. The effect of the insulin/IGF pathway on GFPT2 has not been reported before. Phosphoproteomics



comparison of D492 and D492HER2 also confirmed altered downstream signaling within the ERK/MAPK pathway. Kinase enrichment analysis highlighted ERK/MAPK and components of the Wnt signaling pathway, *i.e.*, GSK3- $\beta$  and casein kinase. We interrogated GSK3- $\beta$  on account of its role in responding to oxidative stress (85–88) and regulation of Wnt signaling of importance to the EMT program (89). GSK3- $\beta$  phosphorylation was higher in D492HER2 suggesting inactivation of GSK3- $\beta$  *via* phosphorylation (90). GSK3- $\beta$  RNA and protein levels were increased in D492, indicating more active GSK3- $\beta$  kinase activity in D492. Knockdown of GSK3- $\beta$  in the D492 cells upregulated GFPT2 expression suggesting that GSK3- $\beta$  suppresses GFPT2. GSK3 kinases are regulators of multiple complex biological functions, which can be inhibited by insulin/IGF1 and ERK/MAPK (88, 91–93). Our data suggest that GFPT2 is under the control of EGF and insulin and down-regulated in epithelial D492 by GSK3- $\beta$  responding to oxidative stress.

In summary (Fig. 9), GFPT2 is upregulated in breast EMT with higher expression in partial EMT, and knockdown of GFPT2 in mesenchymal cells decreases the EMT marker VIM and cell proliferation and invasion as well as HBP flux. It is a marker for claudin-low breast cancers and oxidative stress. GFPT2 is under the regulation of the insulin and EGF signaling networks within which GSK3- $\beta$  imparts a negative effect. Further studies are needed to understand the mechanism of GFPT2 regulating cystathionine, SQOR, the role of GFPT2 in oxidative stress and the transsulfuration pathway, and to confirm the effects of insulin, EGF, and GSK- $\beta$  on GFPT2 in more cell lines.

#### DATA AVAILABILITY

The mass spectrometry proteomics data (LFQ) have been deposited to the ProteomeXchange Consortium *via* the PRIDE (94) partner repository with the dataset identifier PXD025600. The MS/MS spectra have been deposited on MS-Viewer with search keys kkpzmzoav00 (LFQ) and ijjdvuqz6w (iBAQ).

The mass spectrometry proteomics data (SILAC) have been deposited to the ProteomeXchange Consortium *via* the PRIDE (94) partner repository with the dataset identifier PXD025858. The MS/MS spectra have been deposited on MS-Viewer with search keys uwpnz5jwk0 (replicate 1), jnuipmemjk (replicate 2), and ztgmyvh5le (replicate 3) and fhlawvolf (iBAQ).

**Supplemental data**—This article contains [supplemental data](#) (43).

**Acknowledgments**—We are grateful for the contributions from Douglas Lamont and Amy Tavendale at the “Finger-Prints” Proteomics Facility, College of Life Sciences, MSI/WTB/JBC Complex, University of Dundee, for their help with the proteomic and phosphoproteomic data collection and analysis. The graphical abstract was created with [BioRender](#).

**com.** This work was funded by Icelandic Center for Research (RANNÍS, 163254-052) and Góngum Saman 2018.

**Author contributions**—Q. W. and O. R. conceptualization; Q. W. and S. T. K. formal analysis; O. R. funding acquisition; Q. W. investigation; Q. W., S. T. K., and A. K. methodology; O. R. project administration; T. G. and O. R. resources; S. H. and O. R. supervision; Q. W. validation; Q. W. visualization; Q. W. and O. R. writing—original draft; Q. W., S. T. K., A. K., T. G., S. H., and O. R. writing—review and editing.

**Conflict of interest**—The authors declare no competing interests.

**Abbreviations**—The abbreviations used are: AKAP12, A-kinase anchor protein 12; AKR1B1, Aldo-keto reductase family 1, member B1 (aldose reductase), isoform CRA\_a; AKR1C1, Aldo-keto reductase family 1, member C1; ALDH1A3, Aldehyde dehydrogenase 1 family, member A3, isoform CRA\_a; ANXA3, Annexin A3; BP, Biological process; BRENCs, Breast endothelial cells; CADM3, Cell adhesion molecule 3; CCLE, Cancer Cell Line Encyclopedia; CDH1, Cadherin-1; CDH2, Cadherin-2; CID, Collision-induced dissociation; CTGF, Connective tissue growth factor; DCD, Dermcidin; DDA, Data-dependent acquisition; DSP, Desmoplakin; DTT, Dithiothreitol; EDTA, Ethylenediaminetetraacetic acid; EGF, Epidermal growth factor; EMT, Epithelial-mesenchymal transition; EPCAM, Epithelial cell adhesion molecule; ERBB2 (HER2), Erb-B2 receptor tyrosine kinase 2; ERK/MAPK, Mitogen-activated protein kinase; FASP, Filter-aided sample preparation; FBS, Fetal bovine serum; FDR, False discovery rate; FGFBP1, Fibroblast growth factor-binding protein 1; FLNC, Filamin-C; GALE, UDP-glucose 4-epimerase; GALNT7, N-acetylgalactosaminyltransferase 7; GES, Gene expression studies; GFPT1, Glutamine-fructose-6-phosphate aminotransferase 1; GFPT2, Glutamine-fructose-6-phosphate transaminase 2; GlcNAc-P, N-acetylglucosamine phosphate; GLUT1, Glucose transporter 1; GLUT4, Glucose transporter type 4; GO, Gene ontology; GSH, Reduced glutathione; GSK3- $\beta$ , Glycogen synthase kinase 3 beta; GSSG, Oxidized glutathione; H<sub>2</sub>O<sub>2</sub>, Hydrogen peroxide; H<sub>2</sub>S, Hydrogen sulfide; HBP, Hexosamine biosynthesis pathway; HER2 (ERBB2), Human epidermal growth factor receptor 2; HMS LINCS, Harvard Medical School Library of Integrated Network-based Cellular Signatures; HPDL, 4-hydroxyphenylpyruvate dioxygenase-like protein; HUVECs, Human umbilical vein endothelial cells; IAA, Iodoacetamide; iBAQ, Intensity-based absolute quantification; IGF, Insulin like growth factor; IGF1R, Insulin like growth factor 1 receptor; IL18, Interleukin-18; IPA, Ingenuity Pathway Analysis; ITGA6, Integrin subunit alpha 6; ITGB4, Integrin subunit alpha 4; K5/6/8/14/19, Keratin 5/6/8/14/19; KEGG, Kyoto Encyclopedia of Genes and Genomes; KRAS, KRAS proto-oncogene, GTPase; KRT1, Keratin 1; LAD1, Ladinin-1; LAMA3, Laminin subunit alpha 3; LAMB3, Laminin subunit beta 3; LFQ, Label-free quantification; LKB1, Serine/threonine kinase 11 (STK11); MGST1, Microsomal glutathione S-transferase 1; MMS, Mesenchymal metabolic

signature; MYL9, Myosin light chain 9; NDRG1, N-Myc downstream regulated 1; NF- $\kappa$ B, Nuclear factor kappa B; NSCLC, Non-small-cell lung cancer; NT5E, 5'-Nucleotidase Ecto; ODC, Ornithine decarboxylase; OGT, O-Linked N-Acetylglucosamine (GlcNAc) Transferase; PBS, Phosphate-buffered saline; PCSK1N, Proprotein convertase subtilisin/kexin type 1 Inhibitor; PFA, Paraformaldehyde; PGM2L1, Glucose 1,6-bisphosphate synthase; PGM3, Phosphoacetylglucosamine mutase; PKC $\alpha$ , Protein kinase C alpha; PKP2, Plakophilin-2; POMC, Pro-opiomelanocortin; PPP, Pentose phosphate pathway; PRSS23, Serine protease 23; PVDF, Polyvinylidene fluoride; PYGB, Glycogen phosphorylase, brain form; RCN3, Reticulocalbin 3; RELA, RELA Proto-Oncogene, NF- $\kappa$ B subunit, transcription factor p65; RT, Room temperature; RTK, Receptor tyrosine kinase; RT-qPCR, Quantitative reverse transcription PCR; S100A14, S100 calcium binding protein A14; S100A2, S100 calcium binding protein A2; SD, Standard deviation; SDS, Sodium dodecyl sulfate; SERPINB5, Serpin family B member 5; SERPINE1, Serpin family E member 1; SILAC, Stable isotope labeling by amino acids in cell culture; SIRT6, Sirtuin 6; SLC2A4, Solute carrier family 2 member 4; SLP-2, Stomatin-like protein 2; SQOR, Sulfide:quinone oxidoreductase; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; sXBP1, Spliced X-box binding protein 1; TAGLN, Transgelin; TCA, Tricarboxylic acid cycle; TCGA, The Cancer Genome Atlas; TFA, Trifluoroacetic acid; TGF- $\beta$ , Transforming growth factor beta; TNF $\alpha$ , Tumor necrosis factor alpha; TWIST, Twist family BHLH transcription factor 1; UAP1, UDP-N-acetylhexosamine pyrophosphorylase; UDP, Uridine diphosphate; UDP-Glc, UDP-glucose; UDP-GlcA, UDP-glucuronate; UDP-GlcNAc, UDP-N-acetylglucosamine; UGDH, UDP-glucose 6-dehydrogenase; UTP, Uridine-5'-triphosphate; VIM, Vimentin.

Received May 26, 2021, and in revised form, October 20, 2021  
Published, MCPRO Papers in Press, December 17, 2021, <https://doi.org/10.1016/j.mcpro.2021.100185>

## REFERENCES

- Britt, K. L., Cuzick, J., and Phillips, K. A. (2020) Key steps for effective breast cancer prevention. *Nat. Rev. Cancer* **20**, 417–436
- Weigelt, B., Peterse, J. L., and van 't Veer, L. J. (2005) Breast cancer metastasis: Markers and models. *Nat. Rev. Cancer* **5**, 591–602
- Hay, E. (1968) Organization and Fine Structure of Epithelium and Mesenchyme in the Developing Chick Embryo. In: Fleischmajer, R., Billingham, R. E., eds. *Epithelial-mesenchymal interactions: Proceedings of the 18th Hahnemann Symposium*, Williams and Wilkins Co., Baltimore: 31–55
- Dongre, A., and Weinberg, R. A. (2019) New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84
- Nieto, M. A., Huang, R. Y., Jackson, R. A., and Thiery, J. P. (2016) EMT: 2016. *Cell* **166**, 21–45
- Lambert, A. W., Pattabiraman, D. R., and Weinberg, R. A. (2017) Emerging biological principles of metastasis. *Cell* **168**, 670–691
- Rivenson-Segal, D., Margalit, R., and Degani, H. (2002) Glycolysis as a metabolic marker in orthotopic breast cancer, monitored by *in vivo* (13)C MRS. *Am. J. Physiol. Endocrinol. Metab.* **283**, E623–E630
- Barnabas, G. D., Lee, J. S., Shami, T., Harel, M., Beck, L., Selltreinik, M., Jerby-Amon, L., Erez, N., Ruppin, E., and Geiger, T. (2021) Serine biosynthesis is a metabolic vulnerability in IDH2-driven breast cancer progression. *Cancer Res.* **81**, 1443–1456
- Halldorsson, S., Rohatgi, N., Magnusdottir, M., Choudhary, K. S., Gudjonsson, T., Krutusen, E., Barkovskaya, A., Hilmarsdottir, B., Perander, M., Maelandsmo, G. M., Gudmundsson, S., and Rolfsson, Ó. (2017) Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett.* **396**, 117–129
- Lucenay, K. S., Doostan, I., Karakas, C., Bui, T., Ding, Z., Mills, G. B., Hunt, K. K., and Koyomarsi, K. (2016) Cyclin E associates with the lipogenic enzyme ATP-citrate lyase to enable malignant growth of breast cancer cells. *Cancer Res.* **76**, 2406–2418
- Ella, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., Verfaillie, C., Grünwald, T. G. P., and Fendt, S. M. (2017) Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **8**, 15267
- Lu, Y., Wang, X., Li, X., Xu, G., Bai, Y., Wu, J., Piao, Y., Shi, Y., Xiang, R., and Wang, L. (2020) Nucleotide de novo synthesis increases breast cancer stemness and metastasis via cGMP-PKG-MAPK signaling pathway. *PLoS Biol.* **18**, e3000872
- Bergers, G., and Fendt, S. M. (2021) The metabolism of cancer cells during metastasis. *Nat. Rev. Cancer* **21**, 162–180
- Inghthorsson, S., Andersen, K., Hilmarsdottir, B., Maelandsmo, G. M., Magnússon, M. K., and Gudjonsson, T. (2016) HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by co-expression of EGFR. *Oncogene* **35**, 4244–4255
- Sigurðsson, V., Hilmarsdottir, B., Sigmundsdottir, H., Fridriksdottir, A. J., Hingnór, M., Villadsen, R., Borg, A., Agnarsson, B. A., Petersen, O. W., Magnússon, M. K., and Gudjonsson, T. (2011) Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One* **6**, e23833
- Briem, E., Inghthorsson, S., Traustadottir, G. A., Hilmarsdottir, B., and Gudjonsson, T. (2019) Application of the D492 cell lines to explore breast morphogenesis, EMT and cancer progression in 3D culture. *J. Mammary Gland Biol. Neoplasia* **24**, 139–147
- Gudjonsson, T., Villadsen, R., Nielsen, H. L., Rønnov-Jessen, L., Bissell, M. J., and Petersen, O. W. (2002) Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev.* **16**, 693–706
- Ackland, M. L., Newgreen, D. F., Fridman, M. C., Waltham, M. C., Arvanitis, A., Minichiello, J., Price, J. T., and Thompson, E. W. (2003) Epidermal growth factor-induced epithelial-mesenchymal transition in human breast carcinoma cells. *Lab. Invest.* **83**, 435–448
- Xu, J., Lamouille, S., and Derynck, R. (2009) TGF- $\beta$ -induced epithelial to mesenchymal transition. *Cell Res.* **19**, 156–172
- Wang, Y., Liu, J., Ying, X., Lin, P. C., and Zhou, B. P. (2016) Twist-mediated epithelial-mesenchymal transition promotes breast tumor cell invasion via inhibition of hippo pathway. *Sci. Rep.* **6**, 24606
- Liu, L., Pan, Y., Ren, X., Zeng, Z., Sun, J., Zhou, K., Liang, Y., Wang, F., Yan, Y., Liao, W., Ding, Y., Liu, X., and Liang, L. (2020) GFPT2 promotes metastasis and forms a positive feedback loop with p65 in colorectal cancer. *Am. J. Cancer Res.* **10**, 2510–2522
- Zhou, L., Luo, M., Cheng, L. J., Li, R. N., Liu, B., and Linghu, H. (2019) Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) promotes the EMT of serous ovarian cancer by activating the hexosamine biosynthetic pathway to increase the nuclear location of  $\beta$ -catenin. *Pathol. Res. Pract.* **215**, 152681
- Karvelsson, S. T., Sigurðsson, A., Seip, K., Grinde, M. T., Wang, Q., Johannsson, F., Maelandsmo, G. M., Moestue, S. A., Rolfsson, Ó., and Halldorsson, S. (2021) EMT-derived alterations in glutamine metabolism sensitize mesenchymal breast cells to mTOR inhibition. *Mol. Cancer Res.* **19**, 1546–1558
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362
- Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021
- Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
- Cox, J., Neuhauser, N., Michalski, A., Scheltemer, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805
- Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed

- normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526
29. Rolfsson, Ó., Johannsson, F., Magnúsdóttir, M., Paglia, G., Sigurjónsson, Ó. E., Bordbar, A., Pálsson, S., Brynjólfsson, S., Guðmundsson, S., and Pálsson, B. (2017) Mannose and fructose metabolism in red blood cells during cold storage in SAGM. *Transfusion* **57**, 2665–2676
  30. Gu, Z., Elis, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849
  31. Gallili, T. (2015) dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720
  32. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat. Methods* **13**, 731–740
  33. Ulgen, E., Ozisik, O., and Sezeman, O. U. (2019) pathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front. Genet.* **10**, 858
  34. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., et al. (2019) The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503
  35. Szklarczyk, D., Gable, A. L., Lyon, D., Jung, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. V. (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613
  36. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504
  37. Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wizenorowicz, M., Sanchez-Vega, F., et al. (2018) Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6
  38. Zhao, M., Kong, L., Liu, Y., and Qu, H. (2015) dbEMT: An epithelial-mesenchymal transition associated gene resource. *Sci. Rep.* **5**, 11459
  39. Ghandi, M., Huang, F. W., Jane-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., 3rd, Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., et al. (2019) Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508
  40. Koleli, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., Vidovic, D., Forlin, M., Kelley, T. T., D'Urso, A., Allen, B. K., Torre, D., Jagodnik, K. M., Wang, L., Jenkins, S. L., et al. (2017) Data portal for the library of integrated network-based cellular signatures (LINCS) program: Integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.* **46**, D558–D566
  41. The Cancer Genome Atlas Network. (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70
  42. Gröger, C. J., Grubinger, M., Waldhör, T., Vierlinger, K., and Mikulits, W. (2012) Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. *PLoS One* **7**, e51136
  43. Zhao, M., Liu, Y., Zheng, C., and Qu, H. (2019) dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and recalculated regulation information for cancer metastasis. *J. Genet. Genomics* **46**, 595–597
  44. Shaul, Y. D., Freinkman, E., Comb, W. C., Cantor, J. R., Tam, W. L., Thiru, P., Kim, D., Kanarek, N., Pacold, M. E., Chen, W. W., Brierie, B., Possemato, R., Reinhardt, F., Weinberg, R. A., Yaffe, M. B., et al. (2014) Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **158**, 1094–1109
  45. Lawrence, R. T., Perez, E. M., Hernandez, D., Miller, C. P., Haas, K. M., Irie, H. Y., Lee, S. I., Blau, C. A., and Villón, J. (2015) The proteomic landscape of triple-negative breast cancer. *Cell Rep.* **11**, 630–644
  46. Loh, C. Y., Chai, J. Y., Tang, T. F., Wong, W. F., Sethi, G., Shanmugam, M. K., Chong, P. P., and Looi, C. Y. (2019) The E-cadherin and N-cadherin switch in epithelial-to-mesenchymal transition: Signaling, therapeutic implications, and challenges. *Cells* **8**, 1118
  47. Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., Garcia de Herreros, A., et al. (2020) Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **21**, 341–352
  48. Oki, T., Yamazaki, K., Kuromitsu, J., Okada, M., and Tanaka, I. (1999) cDNA cloning and mapping of a novel subtype of glutamine:fructose-6-phosphate amidotransferase (GFAT2) in human and mouse. *Genomics* **57**, 227–234
  49. Chen, P. H., Chi, J. T., and Boyce, M. (2018) Functional crosstalk among oxidative stress and O-GlcNAc signaling pathways. *Glycobiology* **28**, 556–564
  50. Simpson, N. E., Tryndyak, V. P., Beland, F. A., and Pogribny, I. P. (2012) An *in vitro* investigation of metabolically sensitive biomarkers in breast cancer progression. *Breast Cancer Res. Treat.* **133**, 959–968
  51. Zitzler, J., Link, D., Schafer, R., Liebetrau, W., Kazinski, M., Bonin-Debs, A., Behl, C., Buckel, P., and Brinkmann, U. (2004) High-throughput functional genomics identifies genes that ameliorate toxicity due to oxidative stress in neuronal HT-22 cells: GFPT2 protects cells against peroxide. *Mol. Cell. Proteomics* **3**, 834–840
  52. Morgan, M. J., and Liu, Z. G. (2011) Crosstalk of reactive oxygen species and NF- $\kappa$ B signaling. *Cell Res.* **21**, 103–115
  53. Szymura, S. J., Zaemes, J. P., Allison, D. F., Clift, S. H., D'Innocenzi, J. M., Gray, L. G., McKenna, B. D., Morris, B. B., Bekiranov, S., LeGallo, R. D., Jones, D. R., and Mayo, M. W. (2019) NF- $\kappa$ B upregulates glutamine-fructose-6-phosphate transaminase 2 to promote migration in non-small cell lung cancer. *Cell Commun. Signal.* **17**, 24
  54. Steinhilber, S. S., Morera, E., Budkova, Z., Schepsky, A., Wang, Q., Rolfsson, Ó., Riedel, A., Krueger, A., Hilmarsdóttir, B., Maalandsmo, G. M., Valdimarsdóttir, B., Sigurdardóttir, A. K., Agnarsson, B. A., Jonasson, J. G., Ingthorsson, S., et al. (2020) ECM1 secreted by HER2-overexpressing breast cancer cells promotes formation of a vascular niche accelerating cancer cell migration and invasion. *Lab. Invest.* **100**, 928–944
  55. Zhang, W., Bouchard, G., Yu, A., Shafiq, M., Jamali, M., Shrager, J. B., Ayers, K., Bakr, S., Gentles, A. J., Diehn, M., Quon, A., West, R. B., Nair, V., van de Rijn, M., Napel, S., et al. (2018) GFPT2-expressing cancer-associated fibroblasts mediate metabolic reprogramming in human lung adenocarcinoma. *Cancer Res.* **78**, 3456–3457
  56. Richani, D., Sutton-McDowall, M. L., Frank, L. A., Gilchrist, R. B., and Thompson, J. G. (2014) Effect of epidermal growth factor-like peptides on the metabolism of *in vitro*-matured mouse oocytes and cumulus cells. *Biol. Reprod.* **90**, 49
  57. Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Jr., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., and Brooks, S. C. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **50**, 6075–6086
  58. Qu, Y., Han, B., Yu, Y., Yao, W., Bose, S., Kartan, B. Y., Giuliano, A. E., and Cui, X. (2015) Evaluation of MCF10A as a reliable model for normal human mammary epithelial cells. *PLoS One* **10**, e0131285
  59. Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527
  60. Bhat-Nakshatri, P., Appaiah, H., Ballas, C., Pick-Franke, P., Goulet, R., Jr., Badve, S., Srour, E. F., and Nakshatri, H. (2010) SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype. *BMC Cancer* **10**, 411
  61. Cailleau, R., Olivé, M., and Cruciger, Q. V. (1978) Long-term human breast carcinoma cell lines of metastatic origin: Preliminary characterization. *In Vitro* **14**, 911–915
  62. Hugo, H. J., Gunasinghe, N. P. A. D., Hollier, B. G., Tanaka, T., Blick, T., Toh, A., Hill, P., Gilles, C., Waltham, M., and Thompson, E. W. (2017) Epithelial requirement for *in vitro* proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: Oncogenic rather than tumor-suppressive role of E-cadherin. *Breast Cancer Res.* **19**, 86
  63. Young, R. K., Cailleau, R. M., Mackay, B., and Reeves, W. J. (1974) Establishment of epithelial cell line MDA-MB-157 from metastatic pleural effusion of human breast carcinoma. *In Vitro* **9**, 239–245



64. Fougner, C., Bergholtz, H., Norum, J. H., and Sorlie, T. (2020) Re-definition of claudin-low as a breast cancer phenotype. *Nat. Commun.* **11**, 1787
65. Li, M., Bu, X., Cai, B., Liang, P., Li, K., Qu, X., and Shen, L. (2019) Biological role of metabolic reprogramming of cancer cells during epithelial-mesenchymal transition (review). *Oncol. Rep.* **41**, 727–741
66. Chiaradonna, F., Ricciardiello, F., and Palorini, R. (2018) The nutrient-sensing hexosamine biosynthetic pathway as the hub of cancer metabolic rewiring. *Cells* **7**, 53
67. Taparra, K., Tran, P. T., and Zachara, N. E. (2016) Hijacking the hexosamine biosynthetic pathway to promote EMT-mediated neoplastic phenotypes. *Front. Oncol.* **6**, 85
68. Gu, Y., Mi, W., Ge, Y., Liu, H., Fan, Q., Han, C., Yang, J., Han, F., Lu, X., and Yu, W. (2010) GlcNAcylation plays an essential role in breast cancer metastasis. *Cancer Res.* **70**, 6344–6351
69. Askari, M., Kordi-Tamandani, D. M., Almadani, N., McElreavey, K., and Totonchi, M. (2019) Identification of a homozygous GFPT2 variant in a family with asthenozoospermia. *Gene* **699**, 16–23
70. Chao, D., Ariake, K., Sato, S., Ohtsuka, H., Takadate, T., Ishida, M., Masuda, K., Maeda, S., Miura, T., Mitachi, K., Yu, X. J., Fujishima, F., Mizuma, M., Nakagawa, K., Morikawa, T., et al. (2021) Stomatin-like protein 2 induces metastasis by regulating the expression of a rate-limiting enzyme of the hexosamine biosynthetic pathway in pancreatic cancer. *Oncol. Rep.* **45**, 90
71. Pommier, R. M., Sanlaville, A., Toton, L., Kielbassa, J., Thomas, E., Ferrari, A., Sertier, A. S., Hollande, F., Martinez, P., Tissier, A., Morel, A. P., Ouzounova, M., and Puisieux, A. (2020) Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat. Commun.* **11**, 3431
72. Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X., and Perou, C. M. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68
73. Kim, H. S., Mendiarrata, S., Kim, J., Pecot, C. V., Larsen, J. E., Zubovych, I., Seo, B. Y., Kim, J., Eskioçak, B., Chung, H., McMillan, E., Wu, S., De Brabander, J., Komurov, K., Toombs, J. E., et al. (2013) Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer. *Cell* **155**, 552–566
74. Kim, J., Lee, H. M., Cai, F., Ko, B., Yang, C., Lieu, E. L., Muhammad, N., Rhyne, S., Li, K., Haloul, M., Gu, W., Faubert, B., Kaushik, A. K., Cai, L., Kashi, S., et al. (2020) The hexosamine biosynthesis pathway is a targetable liability in KRAS/LKB1 mutant lung cancer. *Nat. Metab.* **2**, 1401–1412
75. Eriksson, F. F., Rolfsson, O., Ogmundsdottir, H. M., Haraldsson, G. G., Thorsteinsdottir, M., and Halldorsson, S. (2018) Altered plasmalogen content and fatty acid saturation following epithelial to mesenchymal transition in breast epithelial cell lines. *Int. J. Biochem. Cell Biol.* **103**, 99–104
76. Giannoni, E., Parri, M., and Chiarugi, P. (2012) EMT and oxidative stress: A bidirectional interplay affecting tumor malignancy. *Antioxid. Redox Signal.* **16**, 1248–1263
77. Jezierska-Drutel, A., Rosenzweig, S. A., and Neumann, C. A. (2013) Role of oxidative stress and the microenvironment in breast cancer development and progression. *Adv. Cancer Res.* **119**, 107–125
78. Huber, M. A., Azoitei, N., Baumann, B., Grunert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H., and Wirth, T. (2004) NF- $\kappa$ B is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J. Clin. Invest.* **114**, 569–581
79. Ascencio, K., Dilek, N., Augsburg, F., Panagaki, T., Zuhra, K., and Szabo, C. (2021) Pharmacological induction of mesenchymal-epithelial transition via inhibition of H2S biosynthesis and consequent suppression of ACLY activity in colon cancer cells. *Pharmacol. Res.* **165**, 105393
80. Wang, M., Yan, J., Cao, X., Hua, P., and Li, Z. (2020) Hydrogen sulfide modulates epithelial-mesenchymal transition and angiogenesis in non-small cell lung cancer via HIF-1 $\alpha$  activation. *Biochem. Pharmacol.* **172**, 113775
81. Guo, L., Peng, W., Tao, J., Lan, Z., Hei, H., Tian, L., Pan, W., Wang, L., and Zhang, X. (2016) Hydrogen sulfide inhibits transforming growth factor- $\beta$ 1-induced EMT via Wnt/catenin pathway. *PLoS One* **11**, e0147018
82. Sessions, D. T., and Kashatus, D. F. (2021) Mitochondrial dynamics in cancer stem cells. *Cell. Mol. Life Sci.* **78**, 3803–3816
83. Lunetti, P., Di Giacomo, M., Vergara, D., De Domenico, S., Maffia, M., Zera, V., Capobianco, L., and Ferramosca, A. (2019) Metabolic reprogramming in breast cancer results in distinct mitochondrial bioenergetics between luminal and basal subtypes. *FEBS J.* **286**, 688–709
84. Panarsky, R., Crooks, D. R., Lane, A. N., Yang, Y., Cassel, T. A., Fan, T. W., Linehan, W. M., and Moscow, J. A. (2020) Fumarate hydratase-deficient renal cell carcinoma cells respond to asparagine by activation of the unfolded protein response and stimulation of the hexosamine biosynthetic pathway. *Cancer Metab.* **8**, 7
85. Schafer, M., Goodenough, S., Moosmann, B., and Behl, C. (2004) Inhibition of glycogen synthase kinase 3 beta is involved in the resistance to oxidative stress in neuronal HT22 cells. *Brain Res.* **1005**, 84–89
86. Niringiyumukiza, J. D., Cai, H., Chen, L., Li, Y., Wang, L., Zhang, M., Xu, X., and Xiang, W. (2019) Protective properties of glycogen synthase kinase-3 inhibition against doxorubicin-induced oxidative damage to mouse ovarian reserve. *Biomed. Pharmacother.* **116**, 108963
87. Yan, C., Zhang, X., Miao, J., Yuan, H., Liu, E., Liang, T., and Li, Q. (2020) Farnesol directly targets GSK-3 $\beta$  to activate Nrf2-ARE pathway and protect EA.hy926 cells against oxidative stress-induced injuries. *Oxid. Med. Cell Longev.* **2020**, 5967434
88. Rlis, S., Murray, J. B., and O'Connor, R. (2020) IGF-1 signalling regulates mitochondria dynamics and turnover through a conserved GSK-3 $\beta$ -Nrf2-BNIP3 pathway. *Cells* **9**, 147
89. Yook, J. I., Li, X. Y., Ota, I., Hu, C., Kim, H. S., Kim, N. H., Cha, S. Y., Ryu, J. K., Choi, Y. J., Kim, J., Fearon, E. R., and Weiss, S. J. (2006) A Wnt-Axin2-GSK3 $\beta$  cascade regulates Snail1 activity in breast cancer cells. *Nat. Cell Biol.* **8**, 1398–1406
90. Fang, X., Yu, S. X., Lu, Y., Bast, R. C., Jr., Woodgett, J. R., and Mills, G. B. (2000) Phosphorylation and inactivation of glycogen synthase kinase 3 by protein kinase A. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11960–11965
91. Cohen, P., and Frame, S. (2001) The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2**, 769–776
92. Beuret, E., Grieco, S. F., and Jope, R. S. (2015) Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. *Pharmacol. Ther.* **148**, 114–131
93. Ding, Q., Xia, W., Liu, J. C., Yang, J. Y., Lee, D. F., Xia, J., Bartholomeusz, G., Li, Y., Pan, Y., Li, Z., Bargou, R. C., Qin, J., Lai, C. C., Tsai, F. J., Tsai, C. H., et al. (2005) Erk associates with and primes GSK-3 $\beta$  for its inactivation resulting in upregulation of beta-catenin. *Mol. Cell* **19**, 159–170
94. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Ingugianni, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., et al. (2018) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450
95. Smith, S. E., Mellor, P., Ward, A. K., Kendall, S., McDonald, M., Vizeacoumar, F. S., Vizeacoumar, F. J., Napper, S., and Anderson, D. H. (2017) Molecular characterization of breast cancer cell lines through multiple omic approaches. *Breast Cancer Res.* **19**, 65
96. Taparra, K., Wang, H., Malek, R., Lafargue, A., Barbhuiya, M. A., Wang, X., Simons, B. W., Ballew, M., Nugent, K., Groves, J., Williams, R. D., Shiraishi, T., Verdone, J., Yildirim, G., Henry, R., et al. (2019) O-GlcNAcylation is required for mutant KRAS-induced lung tumorigenesis. *J. Clin. Invest.* **129**, 4924–4937
97. Guan, Y., Bhandari, A., Zhang, X., and Wang, O. (2019) Uridine phosphorylase 1 associates to biological and clinical significance in thyroid carcinoma cell lines. *J. Cell Mol. Med.* **23**, 7438–7448
98. Wehbe, M., Soudja, S. M., Mas, A., Chasson, L., Guinard, R., de Tenboscho, C. P., Verdel, G., Van den Eynde, B., and Schmitt-Verhulst, A. M. (2012) Epithelial-mesenchymal-transition-like and TGF $\beta$  pathways associated with autochthonous inflammatory melanoma development in mice. *PLoS One* **7**, e49419
99. Sitolo, B. N., and Mavri-Damelin, D. (2018) Peroxidase is regulated by the epithelial-mesenchymal transition master transcription factor Snail1. *Gene* **646**, 195–202
100. Eckert, M. A., Coscia, F., Chryplewicz, A., Chang, J. W., Hernandez, K. M., Pan, S., Tienda, S. M., Nahotko, D. A., Li, G., Blaženović, I., Lastra, R. R., Curtis, M., Yamada, S. D., Perets, R., McGregor, S. M., et al. (2019) Proteomics reveals NMT1 as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728
101. Lee, J., Kim, H., Lee, J. E., Shin, S. J., Oh, S., Kwon, G., Kim, H., Choi, Y. Y., White, M. A., Paik, S., Cheong, J. H., and Kim, H. S. (2018) Selective Cytotoxicity of the NAMPT Inhibitor FK866 Toward Gastric Cancer Cells With Markers of the Epithelial-Mesenchymal Transition, Due to Loss of NAPRT. *Gastroenterology* **155**, 799–814.e713



## GFPT2 Responses to Oxidative Stress in Mesenchymal Cells

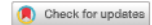
102. Bonet, F., Dueñas, Á., López-Sánchez, C., García-Martínez, V., Aránega, A. E., and Franco, D. (2015) miR-23b and miR-199a impair epithelial-to-mesenchymal transition during atrioventricular endocardial cushion formation. *Dev. Dyn. off. Publ. Am. Assoc. Anatomists* **244**, 1259–1275
103. Alonso, S. R., Tracey, L., Ortiz, P., Pérez-Gómez, B., Palacios, J., Pollán, M., Linares, J., Serrano, S., Sáez-Castillo, A. I., Sánchez, L., Pajares, R., Sánchez-Aguilera, A., Artiga, M. J., Piris, M. A., et al. (2007) A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. *Cancer Res* **67**, 3450–3460
104. Reka, A. K., Chen, G., Jones, R. C., Amunugama, R., Kim, S., Karnovsky, A., Standiford, T. J., Beer, D. G., Omenn, G. S., and Keshamouni, V. G. (2014) Epithelial-mesenchymal transition-associated secretory phenotype predicts survival in lung cancer patients. *Carcinogenesis* **35**, 1292–1300
105. Kwon, Y. J., Baek, H. S., Ye, D. J., Shin, S., Kim, D., and Chun, Y. J. (2016) CYP1B1 enhances cell proliferation and metastasis through induction of EMT and activation of Wnt/beta-Catenin signaling via Sp1 upregulation. *PLoS One* **11**, e0151598
106. Sánchez-Martínez, R., Cruz-Gil, S., Gómez de Cedrón, M., Álvarez-Fernández, M., Vargas, T., Molina, S., García, B., Herranz, J., Moreno-Rubio, J., Reglero, G., Pérez-Moreno, M., Fellu, J., Malumbres, M., and Ramírez de Molina, A. (2015) A link between lipid metabolism and epithelial-mesenchymal transition provides a target for colon cancer therapy. *Oncotarget* **6**, 38719–38736
107. Zhou, C., Nitschke, A. M., Xiong, W., Zhang, Q., Tang, Y., Bloch, M., Elliott, S., Zhu, Y., Bazzone, L., Yu, D., Weldon, C. B., Schiff, R., McLachlan, J. A., Beckman, B. S., Wiese, T. E., et al. (2008) Proteomic analysis of tumor necrosis factor-alpha resistant human breast cancer cells reveals a MEK5/Erk5-mediated epithelial-mesenchymal transition phenotype. *Breast Cancer Res* **10**, R105
108. Lee, J. H., Zhao, X. M., Yoon, I., Lee, J. Y., Kwon, N. H., Wang, Y. Y., Lee, K. M., Lee, M. J., Kim, J., Moon, H. G., In, Y., Hao, J. K., Park, K. M., Noh, D. Y., and Han, W. (2016) Integrative analysis of mutational and transcriptional profiles reveals driver mutations of metastatic breast cancers. *Cell Discov* **2**, 16025–16025
109. Song, S., Yu, W., Lin, S., Zhang, M., Wang, T., Guo, S., and Wang, H. (2018) LncRNA ADPGK-AS1 promotes pancreatic cancer progression through activating ZEB1-mediated epithelial-mesenchymal transition. *Cancer Biol Ther* **19**, 573–583
110. Tang, H., Li, J., Liu, X., Wang, G., Luo, M., and Deng, H. (2016) Down-regulation of HSP60 suppresses the proliferation of glioblastoma cells via the ROS/AMPK/mTOR pathway. *Sci Rep* **6**, 28388
111. Vergara, D., Simeone, P., Latorre, D., Cascione, F., Leporatti, S., Trerotola, M., Giudetti, A. M., Capobianco, L., Lunetti, P., Rizzello, A., Rinaldi, R., Alberti, S., and Maffia, M. (2015) Proteomics analysis of E-cadherin knockdown in epithelial breast cancer cells. *J. Biotechnol.* **202**, 3–11
112. Sun, L., Kong, Y., Cao, M., Zhou, H., Li, H., Cui, Y., Fang, F., Zhang, W., Li, J., Zhu, X., Li, Q., Song, T., and Zhang, T. (2017) Decreased expression of acetyl-CoA synthase 2 promotes metastasis and predicts poor prognosis in hepatocellular carcinoma. *Cancer Sci* **108**, 1338–1346
113. Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T., Choi, H., El Rayes, T., Ryu, S., Troeger, J., Schwabe, R. F., Vahdat, L. T., Altorki, N. K., and Mittal, V. (2015) Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476
114. Tanaka, H., and Ogishima, S. (2015) Network biology approach to epithelial-mesenchymal transition in cancer metastasis: three stage theory. *J. Mol. Cell Biol.* **7**, 253–266
115. Yang, Y., Zhou, X., Xu, M., Piao, J., Zhang, Y., Lin, Z., and Chen, L. (2017) beta-lapachone suppresses tumour progression by inhibiting epithelial-to-mesenchymal transition in NQO1-positive breast cancers. *Sci Rep* **7**, 2681



## Paper III



## ARTICLE OPEN



## Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer

Sigurdur Trausti Karvelsson<sup>1</sup>, Qiong Wang<sup>1</sup>, Bylgja Hilmarsdóttir<sup>2</sup>, Arnar Sigurðsson<sup>3</sup>, Siver Andreas Moestue<sup>4,5</sup>, Gunhild Mari Mælandsdóttir<sup>6</sup>, Skarphedinn Halldórsson<sup>1,6</sup>, Steinn Gudmundsson<sup>1,7</sup> and Ottar Rólfsson<sup>1,8</sup>

Epithelial-to-mesenchymal transition (EMT) is fundamental to both normal tissue development and cancer progression. We hypothesized that EMT plasticity defines a range of metabolic phenotypes and that individual breast epithelial metabolic phenotypes are likely to fall within this phenotypic landscape. To determine EMT metabolic phenotypes, the metabolism of EMT was described within genome-scale metabolic models (GSMMs) using either transcriptomic or proteomic data from the breast epithelial EMT cell culture model D492. The ability of the different data types to describe breast epithelial metabolism was assessed using constraint-based modeling which was subsequently verified using <sup>13</sup>C isotope tracer analysis. The application of proteomic data to GSMMs provided relatively higher accuracy in flux predictions compared to the transcriptomic data. Furthermore, the proteomic GSMMs predicted altered cholesterol metabolism and increased dependency on argininosuccinate lyase (ASL) following EMT which were confirmed in vitro using drug assays and siRNA knockdown experiments. The successful verification of the proteomic GSMMs afforded iBreast2886, a breast GSMM that encompasses the metabolic plasticity of EMT as defined by the D492 EMT cell culture model. Analysis of breast tumor proteomic data using iBreast2886 identified vulnerabilities within arginine metabolism that allowed prognostic discrimination of breast cancer patients on a subtype-specific level. Taken together, we demonstrate that the metabolic reconstruction iBreast2886 formalizes the metabolism of breast epithelial cell development and can be utilized as a tool for the functional interpretation of high throughput clinical data.

npj Systems Biology and Applications (2021)7:36; <https://doi.org/10.1038/s41540-021-00195-5>

## INTRODUCTION

Roughly 90% of all cancer-related deaths are believed to be caused by secondary metastatic tumors<sup>1</sup>. Multiple enzymes have been identified that support cancer cell dissemination in breast cancer through alterations of core metabolic pathways. These include the glycolytic enzymes HK1 and PKM2<sup>2,3</sup>, IDH1 involved in the tricarboxylic acid (TCA) cycle<sup>4</sup>, ACLY in fatty acid synthesis<sup>5</sup>, and PRODH from proline synthesis<sup>6</sup>. Definitive metabolic patterns that differentiate between invasive and non-invasive cancer cells however remain elusive<sup>7</sup>.

One way that epithelial cells gain invasive properties is through the developmental process known as epithelial-to-mesenchymal transition (EMT). When localized breast cancer epithelial cells go through EMT, they gain invasive and apoptosis-resistant properties that contribute to their ability to migrate through the extracellular matrix and form secondary tumors through mesenchymal-to-epithelial transition (MET)<sup>8–10</sup>. Metabolic alterations are believed to be a hallmark of cancer and tumor progression<sup>11</sup> and thus, an overall understanding of the metabolic changes that accompany EMT and MET in breast tissue may help to recognize potential biomarkers and drug targets associated with cancer progression.

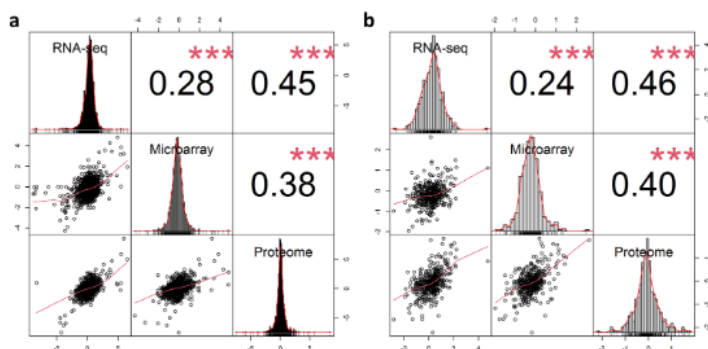
Genome-scale metabolic models (GSMMs) have been successfully used to analyze and interpret changes to cancer metabolism based upon high-throughput datasets<sup>12–14</sup>. GSMM-based studies have revealed significant alterations in the reducing potential during breast tumor development where NADPH is increasingly directed towards reactive oxygen species (ROS) defenses<sup>15</sup>.

Furthermore, the predicted metabolic variability between patients has been utilized successfully for their prognosis<sup>14</sup>. These studies<sup>14,15</sup> were based on transcriptomic or proteomic data obtained from the cell lines or tumors of interest but lacked direct measurements of uptake/secretion rates that constrain metabolic flux as these measurements are challenging to obtain in a clinical setting. Directly incorporating metabolic measurements is expected to provide more accurate predictions than clinical breast cancer data alone.

We hypothesize that GSMMs representing the metabolic plasticity of EMT may help define the metabolism of breast tissue and contribute to the identification of metabolic vulnerabilities for breast cancer diagnostic or therapeutic purposes. The epithelial-derived D492 cell EMT model is comprised of two cell lines (D492 and D492M) that allow metabolic differences that occur following spontaneous EMT in cell culture to be investigated<sup>16</sup>. Similar cell models previously used to study EMT include HMLE and the PMC42 EMT cell models<sup>17–19</sup>.

In order to describe the metabolic plasticity of EMT we recently reported the metabolism of D492 and its mesenchymal-like counterpart D492M by integrated analyses of extracellular metabolomic- and transcriptomic data within tailored GSMMs. The metabolic alterations that occur following EMT in D492<sup>16</sup> mirrored results from a comprehensive analysis of EMT metabolism<sup>20</sup> and anchorage-independent growth<sup>21</sup>. A decrease in glycolysis and changes to mitochondrial oxidation of amino acids, specifically glutamine, threonine, arginine and lysine were observed. Those analyses were limited to transcriptomic and

<sup>1</sup>Center for Systems Biology, University of Iceland, Reykjavik, Iceland. <sup>2</sup>Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. <sup>3</sup>Department of Chemistry, Sekr. TC 2, Faculty II Mathematics and Natural Sciences, Technische Universität Berlin, Berlin, Germany. <sup>4</sup>Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway. <sup>5</sup>Department of Pharmacy, Nord University, Namsos, Norway. <sup>6</sup>Institute for Surgical Research, Vilhelm Magnus Laboratory, Oslo University Hospital, Oslo, Norway. <sup>7</sup>School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. <sup>8</sup>email: ottar@hi.is



**Fig. 1** Correlation of the log-fold differences in D492 and D492M of common gene identifiers between RNA sequencing, microarray and proteomic data. **a** Correlation of log-fold differences of all common gene identifiers ( $n = 2271$ ). **b** Correlation of log-fold differences of common metabolic gene identifiers ( $n = 395$ ). Spearman's rank correlation coefficient was used. The asterisks represent a significant correlation ( $p < 0.001$ ).

extracellular metabolomics data prompting the question of how proteomic data would alter the predictions of D492 metabolic network activity given the nonlinear relationship of transcription and translation<sup>22,23</sup>.

Here, we extend the D492 EMT GSMM, now termed iBreast2886, to include differences in protein levels, further formulating the metabolism of EMT and investigate the models ability to describe breast tissue metabolism. In order to capture the intracellular metabolotypes that accompany EMT in D492 and identify biomarkers that discriminate between the two phenotypes, we used constraint-based modeling and comparative metabolic analysis. In order to reconcile the predicted differences in metabolic phenotypes based on the different data types, we carried out enzyme inhibitor assays,  $1\text{-}^{13}\text{C}$ -glutamine tracer analyses, and siRNA knockdown experiments in vitro to determine the actual phenotypes D492 and D492M cells. Finally, we demonstrate how iBreast2886 can be used as a tool for functional interpretation of tumor gene expression data from breast cancer patients.

## RESULTS

### Direct comparison of different data types reveals their low overlap

In order to determine the consistency of the three different types of data used in this study (microarray, RNA sequencing (RNA-seq) and proteomic) for D492 epithelial cells and D492M mesenchymal cells, we compared the three data types by calculating the Spearman correlation of the log-fold differences between D492M and D492 (Fig. 1a).

The correlation between RNA-seq and proteomic data was the highest ( $\rho = 0.46$ ) and the correlation between the two gene expression methods was lowest ( $\rho = 0.28$ ). By comparing only the metabolic identifiers, the correlation between the dataset did not change (Fig. 1b).

To compare the datasets on a more functional metabolic level, we investigated and compared their ability to infer metabolic activity of D492 and D492M using constraint-based metabolic modeling<sup>24</sup>. In order to achieve this, we used the different datasets as constraints on our previous reconstruction of breast metabolism, which we refer to hereafter as iBreast2886.

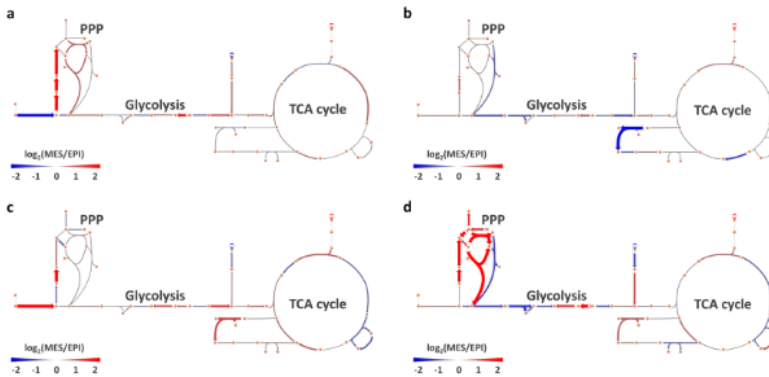
### True metabolic flux is reflected in cell-specific metabolic networks from proteomic data but not other data types

For the comparative metabolic analysis, we constructed GSMMs based on RNA-seq and proteomic data from the epithelial D492 and mesenchymal D492M and compared these to microarray-based GSMMs built previously<sup>16</sup> and iBreast2886 GSMMs where only the extracellular constraints were applied. Henceforth, these will be referred to as the RNA-seq GSMMs, protein GSMMs, microarray GSMMs, and media GSMMs.

In order to compare the EPI and MES models in all pairs of GSMMs, representative flux values (flux profiles) for all reactions that obey the GSMM steady-state assumptions for all models were obtained through random sampling of the solution space<sup>25</sup>, where the median values for reactions were used to represent their activity. The relative differences between EPI and MES in all four GSMM pairs are summarized in Fig. 2a–d. Hierarchical clustering of the GSMMs flux profiles revealed highest similarity between the RNA-seq- and proteomic-constrained models on a phenotype-specific level (Supplementary figure 1). Upon closer inspection, it was clear that reaction similarity was different in the various subcellular compartments. Specifically, the flux similarity of the RNA-seq- and proteomic-constrained models was compartment specific, where the endoplasmic reticulum (ER) had the highest correlation of EMT-linked differences in reaction activity, followed by the cytosol and mitochondria (Supplementary table 1).

As the ground truth for the comparative analysis of pathway activity within our GSMMs, we used isotope labeling patterns from  $1\text{-}^{13}\text{C}$ -labeled glutamine experiments. This tracer is capable of quantifying the contribution of glutamine, one of two major carbon sources of D492 and D492M<sup>16</sup>, to citrate, malate, and aspartate through reductive carboxylation (Fig. 3a) and to the synthesis of proline and glutathione. The contributions from glutamine to the aforementioned metabolites are not whole metabolic pathways but subsets of reductive glutaminolysis and will be referred to as metabolic routes. Some of these metabolic routes occur in more than one cellular compartment. Reductive glutaminolysis is therefore a good representation of the compartment-based complexity of eukaryotic cellular metabolism.

It is challenging to infer metabolic pathway activity by observing multiple, individual reactions (cf Fig. 2). Therefore, we utilized an activity measure that quantifies metabolic route activity in compartmentalized GSMMs based on random sampling results. From the metabolic route activity calculations, we observed that the different omics-constrained GSMMs had different predictions



**Fig. 2** Relative differences in reaction activity in EPI and MES models constrained in four different ways. **a** Only extracellular constraints, **(b)** Microarray, **(c)** RNA-seq and **(d)** Proteomic data. The pathways shown are glycolysis, TCA cycle, and pentose-phosphate pathway. Red represents higher activity in MES whereas blue represents higher activity in EPI, represented by log-fold differences in median activity from random sampling of the models.

of the production of metabolites derived from glutamine (Fig. 3b). According to the  $1\text{-}^{13}\text{C}$ -labeled glutamine results, there was relatively higher citrate derived from glutamine in D492M than D492, indicating increased flow through reductive carboxylation of glutamine and/or the decreased condensation of oxaloacetate and acetyl-CoA (Fig. 3a). The proteomic GSMMs were the only ones to predict both routes correctly (Fig. 3c and Supplementary figure 2) along with the other four routes that were investigated. The microarray-constrained GSMMs correctly predicted the relative difference in metabolic route activity for only two routes, the RNA-seq GSMMs for three routes, and the GSMMs constrained only with extracellular uptake and secretion rates predicted correctly for four routes in total. Thus, the results indicated the relatively higher validity of the proteomics-constrained GSMMs compared to the other data types for intracellular, compartmentalized flux predictions.

#### Results from GSMMs constrained with proteomic data reveal metabolic vulnerabilities of EMT

For the investigation of EMT-specific metabolic remodeling, we utilized the same methodology as before<sup>16</sup> to identify reactions whose activity specifically requires alteration in order to switch from an epithelial flux profile (EPI) to a mesenchymal one (MES). As the proteomics-constrained GSMMs had the most accurate flux predictions, we used them for this analysis. Briefly, we used a hypergeometric test to identify whether the altered reactions were enriched with any subsystems (e.g. the metabolic pathway families with specific functional roles) within iBreast2886. The results showed that two out of the top four enriched reaction sets among EMT-linked reactions are within cholesterol metabolism (highlighted in red in Fig. 4a).

Statins are a class of drugs that are broadly prescribed to patients with hypercholesterolemia. They work by inhibiting HMG-CoA reductase (Fig. 4b), the rate-limiting step in cholesterol synthesis<sup>26</sup>. We treated the D492 and D492M cells with lovastatin and found that following the successful inhibition of cholesterol synthesis in both cell lines (Fig. 4c), it was apparent that the D492 cells were more sensitive to the drug in terms of survival (Fig. 4d).

Thus, in addition to being the most accurate model in terms of intracellular fluxes, the analysis of the proteomic iBreast2886 GSMMs proved useful in identifying the differences in cholesterol metabolism in D492 and D492M. On the same note, we performed

gene essentiality analysis of the proteomic GSMMs and found the essential genes for EPI and MES, respectively. Focusing particularly on the mesenchymal GSMM due to its metastatic involvement, we found that there were nine genes essential for the MES model. These were Argininosuccinate Lyase (ASL), Ornithine Aminotransferase (OAT), Pyruvate Dehydrogenase Complex Component X (PDHX), Proline Dehydrogenase 1 (PRODH), Renin binding protein (RENBP), Isocitrate Dehydrogenase 2 (IDH2), Guanylate Kinase 1 (GUK1), 6-Phosphogluconolactonase (PGLS) and Cystathionine Gamma-Lyase (CTH).

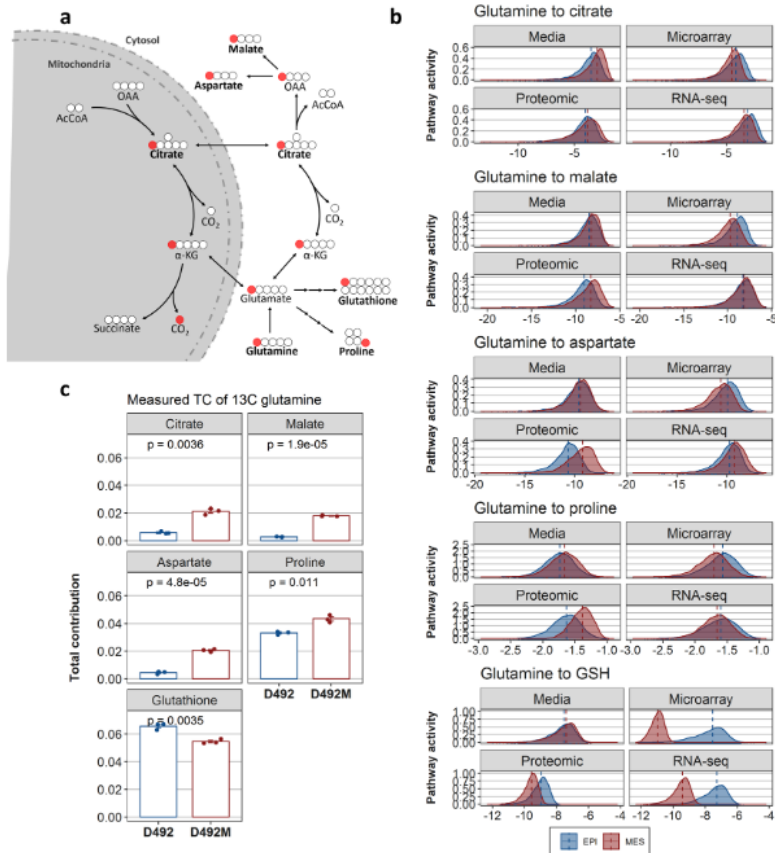
In order to narrow down the list of genes to verify *in vitro*, we evaluated the genes' relationship to survival of patients with claudin-low breast cancer, which is representative for the mesenchymal-like phenotype of breast cancer that expresses high levels of EMT markers<sup>27</sup>. This we achieved by measuring the concordance index (C-index) for the genes, which is a metric for predictive ability of survival models based on gene expression levels<sup>28</sup>. ASL had the highest C-index (and lowest associated *p* value) among the genes (Fig. 5a) and was chosen for *in vitro* survival analysis.

After lowering ASL expression by 75% using small interfering RNAs (siRNA), the survival of D492M cells was significantly diminished whereas the survival of D492 cells was not altered (Fig. 5b and c). The same effect was observed when *GUK1*, the gene with third-lowest *p* value, was silenced in the cells using siRNA (Supplementary figure 3). Importantly, no effect on survival was observed after silencing the gene coding for the neighbor reaction of ASL, argininosuccinate synthase (ASS1) (Supplementary figure 3).

#### iBreast2886-dependent analysis of breast cancer proteome reveals subtype-specific vulnerabilities

Building on the verification of the gene essentiality predictions, we next validated the ability of iBreast2886 to identify growth vulnerabilities in breast cancer that could potentially be exploited for diagnostic or therapeutic purposes. To achieve this, we used proteomic data from breast tumors<sup>29</sup> as constraints for the model. Again, we chose proteomic data (instead of available transcriptomic data) based on our preliminary constraint-based analysis with D492 and D492M data which showed its relatively higher accuracy in capturing intracellular flux phenotypes compared to transcriptomic data.





**Fig. 3** Predicted and measured metabolic route activity in D492 and D492M cells. **a** A carbon tracing map of <sup>13</sup>C-labeled glutamine describing the flow and fate of labeled carbons in the glutamine carbon skeleton. Metabolites in bold are the end metabolites within the pathways we quantified. Red circles represent the <sup>13</sup>C carbon isotopes. The mitochondria is indicated by shaded grey. **b** Density plots of the calculated metabolic route activity (MRA) of five different routes of reductive glutamine metabolism from the total random sampling matrix ( $n = 5800$  flux vectors) for all GSMMs. The blue distributions represent MRA within the epithelial GSMMs whereas red represents MRA within the mesenchymal GSMMs. The dashed line represents the median MRA value. Higher (i.e. more positive) values represent more active routes. All distributions were significantly different in **(b)** ( $p < 0.001$ ) based on a Kolmogorov–Smirnov test. **c** Measured total contribution (TC) from <sup>13</sup>C-glutamine to selected metabolites (after 6 h of cell culture) which represent the same metabolic routes as in **(b)**. Results in **(c)** are shown as mean ± SEM from three experiments (shown with dots). Student's *t*-test was used to estimate significance.

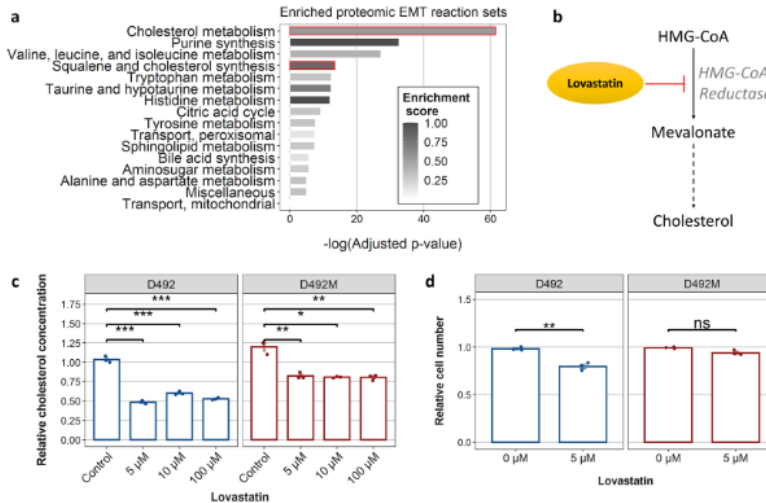
We hypothesized that we would identify subtype-specific metabolic vulnerabilities of breast cancer, i.e. genes specifically essential for estrogen-receptor positive (ER-positive) and ER-negative tumors. The status of the estrogen receptor has repeatedly been shown to be a significant prognostic marker, where patients with ER-negative tumors generally have shorter survival times. After creating 55 patient-specific proteomic GSMMs, we performed gene essentiality analysis on all models. Subsequently, the ER-negative and ER-positive patient GSMMs were tested for over-representation of essential genes using empirical *p* value calculations (see Materials and Methods).

A single essential gene was identified for ER-negative patients: Argininosuccinate Lyase (ASL) (empirical *p* value = 0.0419). In order to confirm these results, we acquired the metadata for the

patients through the Gene Expression Omnibus (GEO) and performed survival analyses using survival in months as time and cancer-related death as event.

Univariate and multivariable Cox proportional hazard models were constructed using age and ASL separately for patients with ER-positive and ER-negative tumors. The Cox analyses (shown in Table 1) revealed that although ASL protein levels were a predictor of death in the univariate models of patients with both ER-positive and ER-negative tumors (HR = 1.16 and 1.12;  $p = 0.067$  and 0.049, respectively), the inclusion of age in the multivariable models attenuated the effects of ASL in the ER-positive patients (HR = 1.08,  $p = 0.44$ ) but not in ER-negative patients (HR = 1.12,  $p = 0.062$ ). Thus, after adjusting for confounding effector age we observed that ASL was a marginally significant predictor of cancer-





**Fig. 4** Integrated analysis of all data-type based GSMMs reveals EMT-related differences in cholesterol metabolism. **a** Enrichment analysis of reactions that need alterations for the EPI to take on a MES flux phenotype in all proteomic GSMMs. This approach helps to identify reaction sets (i.e., families of pathways) which need to be altered for EMT-related changes in flux profiles. Reaction sets shown are ones with FDR-corrected  $p$  value less than 0.01. The scale (Enrichment score) represents the fraction of set reactions within the EMT reactions. The  $p$  values are FDR-adjusted  $p$  values from a hypergeometric test for enrichment of reaction sets. The two reaction sets highlighted in red are cholesterol-related. **b** The mechanism of inhibition of cholesterol synthesis by lovastatin. Dashed arrows represent more than one metabolic reaction. **c** Both D492 and D492M were treated with various concentrations of lovastatin, an inhibitor of cholesterol synthesis. The figure shows the cholesterol concentration in D492 and D492M cells after treatment with lovastatin. **d** Survival of D492 and D492M cells after treatment 5  $\mu\text{M}$  concentration of lovastatin. Results in **c** and **d** are shown as mean  $\pm$  SEM from three experiments (shown with dots). Student's  $t$ -test was used to estimate significance and  $p$  values were adjusted using the Benjamini–Hochberg approach.

related death of only ER-negative patients, which confirms the results from the gene essentiality analysis using breast cancer proteomic data and iBreast2886.

There were no genes significantly enriched for ER-positive patients, likely due to the heterogeneity of this breast cancer subtype which might require further stratification based on the status of the progesterone and HER2 receptors or genes within the PAM50 panel<sup>30</sup>.

## DISCUSSION

High-throughput molecular screening can serve to focus experimental efforts on understanding the functional consequences of molecular variation. Here we have used genome-scale metabolic network modeling to reverse this classification and prioritization strategy. Rather than using high-throughput clinical data as the basis for network analysis of generic metabolic models, we used GSMMs constrained with data from cells in culture whose metabolic phenotypes resemble breast gland development to describe the metabolic landscape of breast epithelium and identify changes in metabolism associated with breast cancer.

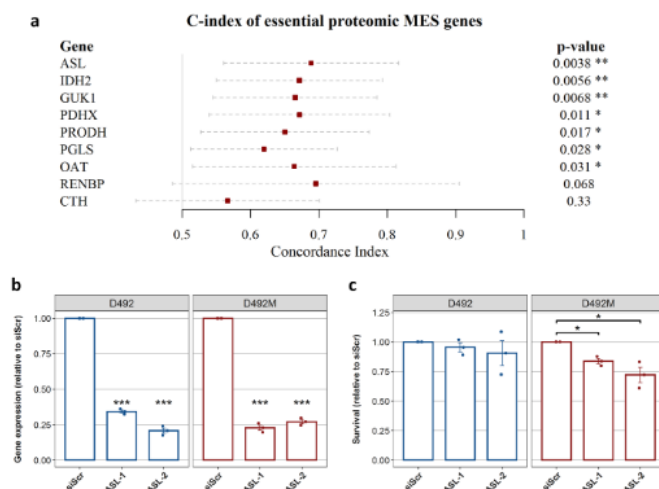
A comparison of the different omics data from the breast epithelial cells D492 and their mesenchymal isogenic cell line D492M revealed a low correlation of the mRNA and protein levels, compatible with literature reports on the correlation of these data types<sup>22,31–33</sup>. There was an even lower correlation of the two different mRNA quantification methods microarray and RNA sequencing (Fig. 1) which is in accordance with the previous studies<sup>34</sup>.

The correlation of transcriptomic and proteomic data can be compartment-specific due to the different spatiotemporal nature

of the molecules<sup>35</sup>. Accordingly, after constraining iBreast2886 with all the different omics data, we found that the differences in fluxes between the proteomic- and RNA-seq-constrained models were indeed highly compartment-specific (Supplementary table 1). Three compartments (cytosol, mitochondria, and the ER) had a significant correlation of  $\rho > 0.37$ , with the ER having the highest value of 0.54. A plausible explanation is that mRNA is synthesized in the nucleus, but is subsequently exported to the rough ER where protein translation takes place. Therefore, the ER displaying high correlation of reaction activity based on proteomic and transcriptomic data is unsurprising<sup>36,37</sup>.

Multiple factors influence the consistency of proteomics and transcriptomics data, not only technical ones like experimental discrepancies and different data-producing platforms, but also biological factors like gene regulation, post-translational modification, different rates of synthesis, and availability of resources<sup>35</sup>. Our findings support that these are different based on the type of cellular compartment and show that the accuracy of GSMMs flux predictions from extracellular uptake and secretion measurements is dependent upon the transcriptomic and proteomic profiles of the cellular compartment of interest.

The compartment-dependent correlation results highlight that care must be taken in the interpretation of metabolic phenotypes from high-throughput data as these may fail to accurately represent the most fundamental parts of energy metabolism. Indeed, it was apparent that the predicted relative activity between EPI and MES was highly data type-dependent (Fig. 2) with the proteomic-constrained GSMMs predicting flux phenotypes most similar to the measured pathway activity using a  $1\text{-}^{13}\text{C}$ -glutamine tracer (Fig. 3). Further analysis of the proteomic GSMMs was successful in proposing valid changes to D492



**Fig. 5 Selection and knockdown of MES-essential genes.** **a** Concordance index (C-index) of proteomic MES essential genes for overall survival prediction of patients with claudin-low breast tumors. Lines represent 95% confidence intervals and *p* values are from the calculation of the C-index using Noether's method<sup>39</sup>. Genes are plotted in descending order based on the *p* values. **b** Expression levels of ASL in D492 and D492M after siRNA-mediated knockdown of the gene. Two different siRNA constructs were used for ASL (ASL-1 and ASL-2) **(c)** siRNA-mediated knockdown of ASL and its effects on the 96 h survival of D492 and D492M. Results in **(b)** and **(c)** are shown as mean ± SEM from three experiments (shown with dots). Student's *t*-test was used to estimate significance and *p* values were adjusted using the Benjamini–Hochberg approach.

**Table 1.** Univariate and multivariable Cox proportional hazard models suggest proteomic levels of ASL are significantly associated with survival of ER-negative breast cancer patients.

ER-status	Variables	Univariate			Multivariable		
		HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value
Positive	ASL	1.16	0.99–1.35	0.067	1.08	0.89–1.31	0.44
	Age	1.07	1.03–1.12	0.0022	1.06	1.02–1.11	0.0086
Negative	ASL	1.12	1.00–1.25	0.049	1.12	0.99–1.25	0.062
	Age	0.99	0.94–1.05	0.77	1.00	0.95–1.05	0.93

The models were created using age (in years) and ASL protein levels. The event used in the survival analysis was cancer-related death.

metabolism following EMT. The EMT-linked reaction list was enriched particularly with reactions taking part in cholesterol and squalene metabolism (Fig. 4a). As a confirmation of these predictions, we found that the cholesterol-inhibiting drug lovastatin had a significantly stronger effect on the survival of D492 than D492M cells (Fig. 4d). Cholesterol has previously been shown to be a promoter of EMT<sup>38</sup> and the cholesterol-inhibiting drug statin has been shown to inhibit cancer invasion and metastasis<sup>39–41</sup>. Importantly, the differences in cholesterol metabolism of D492 and D492M were not captured by a general KEGG pathway enrichment analysis of the significantly different proteins in the cell lines (Supplementary figure 4), suggesting the presence of emergent properties of the iBreast2886 network that are biologically relevant and cannot be elucidated using a generic differential expression analysis.

Similarly, the gene essentiality analysis for the proteomics-constrained MES model suggested that argininosuccinate lyase (ASL) would be essential for D492M which was confirmed by in vitro siRNA knockdown experiments (Fig. 5). Upon knockdown

of the gene, there was a 22.1% reduction in survival of D492M cells on average in contrast to only 7% in D492 cells. This level of survival reduction in D492M is comparable to results from validations in previous studies using similar methodology proposing metabolic targets, where 10–80% reduction in survival have been observed upon in vitro knockdown of the main metabolic target genes<sup>14,42</sup>. A manual investigation of the GSM flux profiles revealed three likely reasons for the essentiality of ASL: 1) compromised proline synthesis via OAT accompanied by 2) decreased fumarate production for the TCA cycle and 3) decreased OAA to aspartate conversion that compromises aminotransferase activity and therefore anaplerotic fueling of the TCA cycle. In addition to ASL, we identified six other significant targets from the gene essentiality analysis, most of which have been associated with poor cancer survival<sup>6,43–45</sup>. For example, increased expression of the *IDH2* gene has been shown to be overexpressed in endometrial, prostate, testicular, and advanced colon cancer<sup>46–48</sup>, and we have recently demonstrated that *IDH2*

indeed fuels reductive glutaminolysis and fatty acid synthesis in the D492 and D492M cells in an accompanying study<sup>19</sup>.

Several studies have carried out metabolic network analysis on a compendium of clinical transcriptomic or proteomic data to extract and prioritize metabolic features of importance<sup>14,15,50–52</sup>. Using our breast metabolic GSMM iBreast2886, we analysed publically available breast tumor proteomic data from 55 breast cancer patients and identified ASL as a metabolic vulnerability of the aggressive ER-negative breast cancer.

Thus, we found that ASL was both essential for D492M cell growth and related to the worse prognosis of ER-negative breast cancer patients. The enzyme coded by ASL, argininosuccinate lyase, produces fumarate and arginine from the breakdown of argininosuccinate. Arginine is a non-essential proteogenic amino acid involved in nitrogen detoxification and the generation of nitric oxide (NO) which is important for invasion and metastasis in various cancer types<sup>6,53</sup>. Downregulation of ASL has been shown to inhibit the growth of breast cancer tumors *in vitro* and *in vivo*<sup>54</sup>. Rabinovich et al.<sup>55</sup> reported that downregulation of *ASS1*, an enzyme directly upstream of ASL, increased pyrimidine synthesis and cancer cell proliferation but did not see the same connection with ASL. The different effect of siRNA knockdown of ASL and *ASS1* on D492 and D492M survival reported here support a mutually exclusive relationship of ASL and *ASS1* as only ASL and not *ASS1* was found to be essential for growth of D492M (Fig. 5c and Supplementary figure 3). This however does not explain the observed differences in the context of linear pathway flux within iBreast2886. A possible explanation is a secondary function of ASL, as it has been shown to influence cyclin A2 levels by direct binding in hepatocellular carcinoma, independent of its enzymatic activity within the *ASS1*-ASL node that also promoted anchorage-independent growth<sup>56</sup>. Intercellular exchange of argininosuccinate between ASL- and *ASS1*-deficient cells, as demonstrated by Davidson et al.<sup>57</sup>, furthermore indicates that the two enzymes need not be co-regulated within a single-cell type. This type of tissue-level metabolic crosstalk would not be captured by our single-cell metabolic reconstruction iBreast2886. Nevertheless, the components of the *ASS1*-ASL node, citrulline, and fumarate, have been reported to be significantly lower and higher, respectively, in ER-negative breast cancer compared to ER-positive which supports altered activity within the *ASS1*-ASL metabolic node<sup>58</sup>. The findings additionally support more studies that have shown that metabolic vulnerabilities of breast cancer lie within arginine metabolism<sup>54,59,60</sup>.

Taken together, the study demonstrates that the metabolism of EMT captured within iBreast2886 is practical for data integration and analysis and that proposed phenotypes are in agreement both with prior investigations of EMT/metastasis and ER-negative breast cancer metabolism. The iBreast2886 reconstruction is first and foremost a metabolic model descriptive of the steady-state metabolic phenotypes that the D492 EMT cell model can achieve based upon the integration of mRNA transcription, protein translation, and metabolite uptake and secretion rates. The integrated analysis of multiple iBreast2886 GSMMs constrained with separate data types collectively yielded more accurate predictions than each individual GSMM, as shown here with the EMT-related changes in cholesterol metabolism and ASL essentiality.

Limitations of iBreast2886 include lack of actual measurements of fatty acid oxidation and cholesterol uptake/secretion rates which might further increase predictive accuracy of iBreast2886. Genes involved in the oxidation of fatty acids are known to correlate with reduced cancer patient survival<sup>4,61,62</sup> and the relationship of cholesterol to EMT and metastasis has been discussed here above<sup>38–41</sup>. The robustness and plasticity of breast tissue metabolism are also more complex than is captured by iBreast2886, which is solely based upon one EMT cell culture model and media constraints that may not accurately reflect the

breast tissue microenvironment<sup>63</sup> and lack flux extremities that may arise from kinetic regulation. Steps towards further understanding of EMT metabolism could be performed by expanding iBreast2886 to account for additional cell lines alongside focused studies aimed at addressing metabolic gaps and network inconsistencies whose presence was demonstrated in this study using isotope tracer analysis. In this way, biochemically accurate descriptions of EMT metabolism in breast tissue to aid in translational cancer research may be pushed forward.

## MATERIALS AND METHODS

### Cell culture

D492 and D492M were cultured in a serum-free H14 medium at 37°C and 5% CO<sub>2</sub> as previously described<sup>19</sup>. H14 is a fully defined medium comprised of DMEM/F12 base with 250 ng/ml insulin, 10 µg/ml transferrin, 10 ng/ml EGF, 2.6 ng/ml sodium selenite, 10<sup>-10</sup> M estradiol, 1.4 µM hydrocortisone, 7.1 ng/ml prolactin, 100 IU penicillin, 0.1 mg/ml streptomycin and 2 mM glutamine. Medium was changed every 48 h while propagating cells, and experiments were performed within four passages. D492 and D492M cells were kindly provided by the Stem Cell Research Unit, University of Iceland, and were screened for *Mycoplasma* infections monthly using PCR-based tests at the Biomedical Center, University of Iceland.

### Origin of iBreast2886 GSMM for breast metabolism

Genome-scale metabolic model construction and analysis were carried out in MATLAB using the COBRA Toolbox<sup>64</sup>. The genome-scale breast tissue metabolic model from Halldorsson et al.<sup>16</sup> was used as a base model. Briefly, RNA sequencing data from both the D492 and D492M cell lines<sup>19</sup> was used to create a metabolic model specific for breast tissue. To achieve this, the human metabolic reconstruction Recon 2 was employed<sup>65</sup>. All genes in the RNA sequencing data with expression values exceeding a fixed cut-off value (1 RPM) in either cell line were identified. To identify the metabolic reactions associated with the list of genes, the Gene-Protein Rules (GPRs) of Recon 2 were used. The FASTCORE model building algorithm<sup>66</sup> was subsequently applied to build a functional metabolic network from the list of identified reactions. The resulting network, referred to as the iBreast2886 model, was manually curated to ensure no metabolites or pathways were blocked or missing.

### Construction and analysis of cell-type-specific epithelial and mesenchymal GSMMs

The iBreast2886 reconstruction was used to create cell-type-specific models of epithelial D492 and mesenchymal D492M based on microarray, proteomic, and RNA sequencing (RNA-seq) data. The workflow of the model construction is outlined in Supplementary figure 5. Briefly, the genes/proteins from each dataset (along with cell-type-specific uptake and secretion rates of multiple metabolites in the media) were used to constrain iBreast2886 to create two models (EPI for epithelial D492, and MES for mesenchymal D492M). Furthermore, the fourth pair of EPI and MES was added that did not contain any intracellular constraints imposed by omics data, but only the cell-type-specific uptake and secretion rates. This gave rise to four pairs of EPI and MES models, where each EPI model shared the same stoichiometry and uptake/secretion rates but had different intracellular reaction constraints based on the different datasets. The same applied to the MES models. See Supplementary information for details.

### Stable isotope tracing analysis

D492 and D492M cells were cultured until confluent as described above. The medium was then changed to a complete H14 containing 1-<sup>13</sup>C-labeled glutamine (Cambridge Isotope Laboratories, Inc., MA, USA). After 6 h of culturing with the <sup>13</sup>C-labeled carbon source, cell metabolism was quenched using cold methanol and intracellular metabolites were extracted using ACN extraction<sup>67</sup>. Analyses were performed on a UPLC system as described in Rolfsson et al.<sup>68</sup>. Results were presented as the total contribution (TC) of carbon sources to measured metabolites<sup>69</sup>.

$$TC = \frac{\sum_{i=0}^n i \cdot m_i}{n} \quad (1)$$



Where  $n$  is the number of C atoms in the metabolite,  $i$  represents the isotopologues and  $m$  is the relative fraction of the isotopologues.

### Comparison of GSMM flux predictions and $^{13}\text{C}$ -labeling profiles

Individual flux distributions from constraint-based modeling of GSMMs were used to estimate the relative contribution of extracellular metabolites to intracellular metabolites of interest. This approach is suitable to measure the flow of carbons between metabolites within GSMMs to ultimately quantify the total activity of specific metabolic routes within pathways that can subsequently be directly compared to definitive results from stable isotope tracer analyses. A schematic explaining the metabolic route activity measure is shown in Supplementary figure 6. A single flux distribution represent the flux values of all reactions within a GSMM which is subject to the constraints applied to the model. In order to calculate the relative flux value  $v_{rel}$  from metabolite  $m_i$  to  $m_{i+1}$  within a pathway of interest, we first identify all consuming reactions of metabolite  $m_i$  using the stoichiometric matrix  $S$ . Then, for a single flux distribution, one can calculate the sum of consuming flux of metabolite  $m_i$ , and the relative flux that is used to produce only metabolite  $m_{i+1}$ , which we will call  $v_{rel}$ . If  $k$  is a consuming reaction of a particular metabolite of interest, then the  $v_{rel}$  value for  $k$  is calculated from the raw flux value of  $k$  divided by the sum of the fluxes of all  $K$  reactions consuming the same metabolite as  $k$ . Therefore, the  $v_{rel}$  of  $k$  (or  $v_{rel}(k)$ ) in a single flux vector is calculated as follows:

$$v_{rel}(k) = \frac{v(k)}{\sum_{i=1}^K v(i)} w_{comp} \quad (2)$$

Where  $w_{comp}$  is the weight given to the relative flux value based on the relative abundance of the compartment it takes place in, since some reactions take place in more than one compartment. The  $v_{rel}(k)$  values for all transport reactions were assumed to be 1. The weights for the compartments were as follows: Cytosol 0.54, mitochondria 0.22, ER 0.12, nucleus 0.06, golgi apparatus 0.03, peroxisomes and lysosomes 0.01, and are representative of their relative volume within cells in general<sup>20</sup>.

Using the relative consumption values for a list of reactions within a single flux vector, it is possible to calculate the metabolic route activity (MRA). To calculate the MRA from metabolite  $m$  to  $m_{i+1}$ , calculate the sum of the log of relative flux values (from Eq. (2)) within that route:

$$MRA = \sum_{i=1}^{k-1} \log(v_{rel}(m_{i+1})) \quad (3)$$

Where the first reaction is the consumption of metabolite  $m_i$  to produce metabolite  $m_{i+1}$ . The MRA of multiple flux vectors (e.g., within a random sampling matrix) can be calculated to get a distribution of MRA within a specific constrained GSMM.

### Lovastatin assay

D492 and D492M cells were treated with 5, 10, and 100  $\mu\text{M}$  concentration of lovastatin (Tocris Bioscience, Bristol, UK) for 24 h after which both cholesterol abundance and cell numbers were assessed. The cholesterol was measured using Amplex<sup>TM</sup> Red Cholesterol Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturers protocol. The cell numbers were assessed using crystal violet staining. Briefly, after 24 h of treatment, the cells were fixed using ice-cold methanol and stained with crystal violet (0.5%). The stain was subsequently released using 10% acetic acid and absorption was measured at 570 nm.

### Scoring of *in silico* gene essentiality candidates

The METABRIC breast cancer clinical dataset<sup>21</sup> was downloaded from cBioPortal<sup>22,23</sup>. The clinical metadata includes information about the claudin-status of the tumors in the data. Using only patients with tumors classified as *claudin-low* and available survival data ( $n=199$ ), we performed a survival analysis. In short, patients were split into two groups based on the best-splitting expression level (as identified through R-package *survminer*'s *surv\_cutpoint*(*t*) function) of a gene of interest. The prognoses of the groups were then examined by calculating the concordance index (C-index)<sup>24</sup>, which provides an overall measure of predictive accuracy of the genes' expression level with right-censored survival data.

### Small interfering RNA (siRNA) knockdown experiments

For the knockdown experiments, Silencer Select siRNAs (Thermo) were used (Negative Control No 1 #4390843), ASL (s1669 and s1671), and ASS1 (s1684). Cells were seeded at density of 3000 cells/well in a 96 well plate. Prior to seeding, the 96 well plates were coated with siRNA and Lipofectamine RNAiMAX (Thermo) for 15 min. Final concentration of siRNA in each well, after addition of cells, was 10 nM. Transfected cells were incubated at 37 °C and 5% CO<sub>2</sub> for 96 h at the end of which cell survival and gene expression were assessed. To measure cell survival, CellTiter Glo Luminescent Cell Viability Assay (Promega, Madison, WI, USA) was used according to instructions of the manufacturer. SpectraMax plate reader was used to measure luminescence at 560 nm. To measure the gene expression, real time quantitative polymerase chain reaction (qPCR) was used.

### Real-time PCR

Total RNA was isolated using TRI-Reagent (Thermo) according to the manufacturer's instructions. RNA concentration was measured using NanoDrop One (Thermo). 0.4 to 1  $\mu\text{g}$  RNA was reverse transcribed to cDNA using High-Capacity cDNA Reverse Transcription kit (Thermo). Real-time quantitative PCR reactions were carried out using Luna Universal qPCR Master Mix (New England Biolabs, Ipswich, MA, USA) according to manufacturer's instructions on a BioRad CFX384 Touch<sup>TM</sup> Real Time System (BioRad Laboratories, Hercules, CA, USA). Gene expression levels were determined with CFX Manager software (BioRad). Primers were designed using the Primer3 software<sup>24</sup>. Primers spanning exon junctions were chosen to ensure specificity. Differences in relative expression were estimated using the  $2^{-\Delta\Delta\text{CT}}$  method. The primer sequences used for quantifying the gene expression were: ASL-fwd 5'-GGAAGCTGTGTTTAA GTGTCA-3', ASL-rev 5'-CAATGTTCTCTGGTGAATCTG-3', ASS1-fwd 5'-CAGG AAGGGGAACGATCAGGT-3', ASS1-rev 5'-CGTGTGCTTTCGTAATCCAT-3', GUK1-fwd 5'-CTTCATCGAGCATGCCGAGTTC-3', GUK1-rev 5'-GAACCTGTATG GCACGAGCAAG-3', ACTB-fwd 5'-CTTCCGGGTGAGTGAGACTG-3' and ACTB-rev 5'-GAGGGAAATGAGGGCAGACTT-3'.

### Analysis of clinical breast cancer data using iBreast2886

Proteomic breast cancer data were acquired from Tang et al.<sup>29</sup>. After removing identifiers with missing data in more than 20% of samples, the data were imported into MATLAB for constraint-based modeling.

Patient-specific GSMMs were constructed from iBreast2886, where the reactions were only constrained in a patient model if their associated protein levels were below the 60th percentile in all patients. The same amount of constraint was applied to the selected reactions as for the EPI and MES models (as described above). The median percentage of constrained reactions in the patients was 3.8%. Gene essentiality analysis was carried out using FBA as described above.

Essential genes that were over-represented in the GSMMs of estrogen receptor (ER) negative ( $n_1=33$ ) and positive patients ( $n_2=32$ ) were identified by randomly sampling  $n_1$  and  $n_2$  patient-specific GSMMs 1000 times from the whole GSMM list. Then, an empirical  $p$  value ( $\hat{p}$ ) was calculated for each gene in the ER-negative and ER-positive patient subsets using the formula from North et al.<sup>24</sup>:

$$\hat{p} = \frac{r+1}{n+1} \quad (4)$$

Where  $\hat{p}$  is the empirical  $p$  value,  $n$  is the number of resampled sets (1000 in this case) and  $r$  is the number of times the resampled sets have an equal or greater number of an essential gene compared to the ER-negative or ER-positive patient sets.

Genes with a  $\hat{p} < 0.05$  were identified and their proteomic levels<sup>29</sup> were tested as subtype-specific survival predictors using the patient metadata acquired from GEO (GSE37751). The metadata used were cancer-related death and survival in months that were acquired using the R-package GEOquery<sup>26</sup>. To assess the effects of genes and confounding variables on patient survival, Cox-proportional hazard models were employed using the R-package survival<sup>27</sup>.

### Statistical analysis

For comparison of two groups, a two-sided Student's  $t$ -test was employed unless when the data did not follow a normal distribution, when the non-parametric Mann-Whitney  $U$ -test was used. When more than a single treatment was compared in the cell lines, the treatments were all compared to the negative control using two-sided Student's  $t$ -test and

subsequently, the  $p$  values were adjusted for multiple comparisons using the Benjamini–Hochberg approach. For the comparison of two distributions (e.g., in the metabolic route activity measurements), a Kolmogorov–Smirnov test was used. Presented data were from at least three independent experiments (represented by dots) and were summarized using mean  $\pm$  standard error. The asterisks in each figure represent the  $p$  values (\* $<0.05$ , \*\* $<0.01$ , \*\*\* $<0.001$ , \*\*\*\* $<0.0001$ , ns = not significant). Statistical methods used for GSMM analysis of breast cancer patients are described in the Analysis of clinical breast cancer data using iBreast2886 section. All statistical analysis was carried out using the R programming language<sup>78</sup>.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. The breast cancer proteomic data that were analysed during the current study are available in Tang et al.<sup>29</sup> and the metadata were acquired from the Gene Expression Omnibus (GEO), accession no. GSE37751.

### CODE AVAILABILITY

All code used in this study for constraint-based modeling and analysis are accessible at [https://github.com/siggitrausti/EMT\\_metabolic\\_modeling](https://github.com/siggitrausti/EMT_metabolic_modeling).

Received: 29 January 2021; Accepted: 9 August 2021;  
Published online: 17 September 2021

### REFERENCES

- Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
- Tseng, P.-L. et al. The decrease of glycolytic enzyme hexokinase 1 accelerates tumor malignancy via deregulating energy metabolism but sensitizes cancer cells to 2-deoxyglucose inhibition. *Oncotarget* **9**, 18949–18969 (2018).
- Wei, L. et al. Leptin promotes epithelial-mesenchymal transition of breast cancer via the upregulation of pyruvate kinase M2. *J. Exp. Clin. Cancer Res.* **35**, 166 (2016).
- Liu, W.-S. et al. Isocitrate dehydrogenase 1-snail axis dysfunction significantly correlates with breast cancer prognosis and regulates cell invasion ability. *Breast Cancer Res.* **20**, 25 (2018).
- Liu, H., Huang, X. & Ye, T. MiR-22 down-regulates the proto-oncogene ATP citrate lyase to inhibit the growth and metastasis of breast cancer. *Am. J. Transl. Res.* **10**, 659–669 (2018).
- Elia, I. et al. Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **8**, 15267 (2017).
- Elia, I., Doglioni, G. & Fendt, S. M. Metabolic hallmarks of metastasis formation. *Trends Cell Biol.* **28**, 673–684 (2018).
- Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420–1428 (2009).
- Ye, X. & Weinberg, R. A. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* **25**, 675–686 (2015).
- Liao, T. T. & Yang, M. H. Revisiting epithelial-mesenchymal transition in cancer metastasis: the connection between epithelial plasticity and stemness. *Mol. Oncol.* **11**, 792–804 (2017).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Mardinoglu, A. & Nielsen, J. Editorial: the impact of systems medicine on human health and disease. *Front Physiol.* **7**, 552 (2016).
- Yizhak, K., Gabay, O., Cohen, H. & Ruppin, E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* **4**, 2632 (2013).
- Yizhak, K. et al. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *eLife* **3**, e03641 (2014).
- Jerby, L. et al. Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res.* **72**, 5712–5720 (2012).
- Halldorsson, S. et al. Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett.* **396**, 117–129 (2017).
- Tam, W. L. et al. Protein kinase C  $\alpha$  is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell* **24**, 347–364 (2013).
- Ackland, M. L. et al. Epidermal growth factor-induced epithelial-mesenchymal transition in human breast carcinoma. *Cells Lab. Investig.* **83**, 435–448 (2003).
- Sigurðsson, V. et al. Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS ONE* **6**, e23833 (2011).
- Shaul, Y. D. et al. Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **158**, 1094–1109 (2014).
- Jiang, L. et al. Reductive carboxylation supports redox homeostasis during anchorage-independent growth. *Nature* **532**, 255–258 (2016).
- Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).
- Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl Acad. Sci. USA* **102**, 2685 (2005).
- Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120 (2014).
- Schellenberger, J. & Palsson, B. O. Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* **284**, 5457–5461 (2009).
- Istvan, E. S. & Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **292**, 1160–1164 (2001).
- Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68–R68 (2010).
- Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA J. Am. Med. Assoc.* **247**, 2543–2546 (1982).
- Tang, W. et al. Integrated proteotranscriptomics of breast cancer reveals globally increased protein-mRNA concordance associated with subtypes and survival. *Genome Med.* **10**, 94 (2018).
- Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- Edfors, F. et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12**, 883 (2016).
- Darmanis, S. et al. Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* **14**, 380–389 (2016).
- Gong, H. et al. Single-cell protein-mRNA correlation analysis enabled by multiplexed dual-analyte co-detection. *Sci. Rep.* **7**, 2776 (2017).
- Trost, B. et al. Concordance between RNA-sequencing data and DNA microarray data in transcriptome analysis of proliferative and quiescent fibroblasts. *R. Soc. Open Sci.* **2**, 150402 (2015).
- Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
- Lu, R. et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**, 358–362 (2009).
- Cooper, G. M. *The Cell - A Molecular Approach* 2nd Edition. (Sunderland (MA): Sinauer Associates, 2000). citeulike-article-id:10266975.
- Alikhani, N. et al. Mammary tumor growth and pulmonary metastasis are enhanced in a hyperlipidemic mouse model. *Oncogene* **32**, 961–967 (2013).
- Taras, D. et al. Pravastatin reduces lung metastasis of rat hepatocellular carcinoma via a coordinated decrease of MMP expression and activity. *J. Hepatol.* **46**, 69–76 (2007).
- Nielsen, S. F., Nordestgaard, B. G. & Bojesen, S. E. Statin use and reduced cancer-related mortality. *N. Engl. J. Med.* **367**, 1792–1802 (2012).
- Zhang, J. et al. Statins, autophagy and cancer metastasis. *Int. J. Biochem. Cell Biol.* **45**, 745–752 (2013).
- Dai, Z. et al. Identification of cancer-associated metabolic vulnerabilities by modeling multi-objective optimality in metabolism. *Cell Commun. Signal.* **17**, 124 (2019).
- Chen, B. et al. MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells. *BMC Cancer* **14**, 443 (2014).
- da Rocha, A. A. et al. Hepatocyte growth factor-regulated tyrosine kinase substrate (HGS) and guanylate kinase 1 (GUK1) are differentially expressed in GH-secreting adenomas. *Pituitary* **9**, 83–92 (2006).
- Liu, Y. et al. Ornithine aminotransferase promoted the proliferation and metastasis of non-small cell lung cancer via upregulation of miR-21. *J. Cell. Physiol.* **234**, 12828–12838 (2019).
- Guirguis, A. et al. Use of gene expression profiles to stage concurrent endometrioid tumors of the endometrium and ovary. *Gynecol. Oncol.* **108**, 370–376 (2008).
- Altenberg, B. & Greulich, K. O. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014–1020 (2004).
- Lv, Q. et al. Altered expression levels of IDH2 are involved in the development of colon cancer. *Exp. Ther. Med.* **4**, 801–806 (2012).
- Trausti Karvelsson, S. et al. *Altered glutamine metabolism exposes EMT derived mesenchymal cells to PI3K/Akt/mTOR pathway inhibition.* (2020) <https://doi.org/10.21203/RS.3.RS-100299/V1>.

50. Björnson, E. et al. Stratification of hepatocellular carcinoma patients based on acetate utilization. *Cell Rep.* **13**, 2014–2026 (2015).
51. Benfeitas, R. et al. Characterization of heterogeneous redox responses in hepatocellular carcinoma patients using network analysis. *EBioMedicine* (2018) <https://doi.org/10.1016/j.ebiom.2018.12.057>.
52. Mardinoglu, A. et al. Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* **9**, (2013).
53. Chen, L.-H., Hsu, W.-L., Tseng, Y.-J., Liu, D.-W. & Weng, C.-F. Involvement of DNMT 3B promotes epithelial-mesenchymal transition and gene expression profile of invasive head and neck squamous cell carcinomas cell lines. *BMC Cancer* **16**, 431 (2016).
54. Huang, H. L. et al. Argininosuccinate lyase is a potential therapeutic target in breast cancer. *Oncol. Rep.* **34**, 3131–3139 (2015).
55. Rabinovich, S. et al. Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* **527**, 379–383 (2015).
56. Hung, Y.-H. et al. Argininosuccinate lyase interacts with cyclin A2 in cytoplasm and modulates growth of liver tumor cells. *Oncol. Rep.* **37**, 969–978 (2017).
57. Davidson, J. S., Baumgarten, I. M. & Harley, E. H. Metabolic cooperation between argininosuccinate synthetase and argininosuccinate lyase deficient human fibroblasts. *Exp. Cell Res* **150**, 367–378 (1984).
58. Budczies, J. et al. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and betalanine metabolism. *J. Proteom.* **94**, 279–288 (2013).
59. Qiu, F. et al. Arginine starvation impairs mitochondrial respiratory function in ASS1-deficient breast cancer cells. *Sci. Signal.* **7**, ra31–ra31 (2014).
60. Cheng, C.-T. et al. Arginine starvation kills tumor cells through aspartate exhaustion and mitochondrial dysfunction. *Commun. Biol.* **1**, 1–15 (2018).
61. Jia, M. et al. Estrogen receptor  $\alpha$  promotes breast cancer by reprogramming choline metabolism. *Cancer Res.* **76**, 5634–5646 (2016).
62. Aiderus, A., Black, M. A. & Dunbier, A. K. Fatty acid oxidation is associated with proliferation and prognosis in breast and other cancers. *BMC Cancer* **18**, 805 (2018).
63. Katzir, R. et al. The landscape of tiered regulation of breast cancer cell metabolism. *Sci. Rep.* **9**, 1–12 (2019).
64. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v3.0. *Nat. Protoc.* **14**, 639–702 (2019).
65. Thiele, I. et al. A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425 (2013).
66. Vlassis, N., Pacheco, M. P. & Sauter, T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.* **10**, e1003424 (2014).
67. Paglia, G. et al. Intracellular metabolite profiling of platelets: evaluation of extraction processes and chromatographic strategies. *J. Chromatogr. B* **898**, 111–120 (2012).
68. Rolfsson, Ö. et al. Mannose and fructose metabolism in red blood cells during cold storage in SAGM. *Transfusion* **57**, 2665–2676 (2017).
69. Buescher, J. M. et al. A roadmap for interpreting <sup>13</sup>C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.* **34**, 189–201 (2015).
70. Alberts, B. et al. *Essential cell biology*. (Garland Science, 2013).
71. Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).
72. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Disco.* **2**, 401–404 (2012).
73. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pii (2013).
74. Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).
75. North, B. V., Curtis, D. & Sham, P. C. A note on the calculation of empirical P values from Monte Carlo procedures [1]. *Am. J. Hum. Genet.* **71**, 439–441 (2002).
76. Sean, D. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
77. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. (Springer New York, 2000). <https://doi.org/10.1007/978-1-4757-3294-8>.
78. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
79. Penciana, M. J. & D'Agostino, R. B. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat. Med.* **23**, 2109–2123 (2004).

## ACKNOWLEDGEMENTS

This work was supported by the Icelandic Research Fund (#163254-051), Göngum Saman, and the Norwegian Research Council (#239940). The authors thank Frey Johannsson and Sarah McGarrity for valuable input regarding constraint-based modeling methodology, <sup>13</sup>C isotope tracer, and metabolomics analysis.

## AUTHOR CONTRIBUTIONS

Conceptualization, S.T.K., S.G. and O.R.; Methodology, S.T.K., A.R., Q.W., S.A.M., S.H., S.G. and O.R.; Software, S.T.K. and S.G.; Formal analysis and Investigation, S.T.K., A.S., Q.W., S.A.M.; Data curation, S.T.K., Q.W., A.S. and S.H.; Visualization, S.T.K.; Writing – Original Draft, S.T.K. and O.R.; Writing—Review and Editing, all authors; Funding acquisition, G.M.M., S.A.M., S.G. and O.R., Supervision, G.M.M., S.A.M., S.G. and O.R.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41540-021-00195-5>.

**Correspondence** and requests for materials should be addressed to Ottar Rolfsson.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## Appendix

**Appendix Table 1.** The topmost dysregulated phosphorylation sites by comparing D492M with D492 after statistical analysis of the SILAC phosphoproteomic dataset (fold change  $\geq 2$ , one-sample Student's T-test p value  $< 0.05$ ;  $p > 0.05$  and fold change  $\geq 2$  in all three replicates). Blue: active in D492; Red: active in D492M. NA: gene name is not available.

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median Log2 (D492/D492M)	p Value
COL17A1_pS148	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	5.341	0.019
CRYBG1_pS280	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	3.123	0.004
COL17A1_pS93	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	3.058	0.108
FAM83B_pS869	Q5T0W9	Protein FAM83B	3.009	0.006
AKAP12_pS1395	Q02952	A-kinase anchor protein 12	2.750	0.000
LMNB2_pS424	Q03252	Lamin-B2	2.642	0.001
CRYBG1_pS299	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	2.587	0.002
TNS4_pS350	Q8LZW8	Tensin-4	2.365	0.100
LRBA_pS2485	P50851	Isoform 2 of Lipopolysaccharide-responsive and beige-like anchor protein	2.345	0.035
NA_pS31	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	2.270	0.002
FAM62A_pS620	Q9BSJ8	Family with sequence similarity 62 (C2 domain containing), member A, isoform CRA_b	2.179	0.003
AKAP12_pS96	Q02952	A-kinase anchor protein 12	2.167	0.027
NA_pS10	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	2.138	0.009
COL17A1_pS85	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	2.116	0.007
AKAP12_pS75	Q02952	A-kinase anchor protein 12	2.061	0.005
EHD2_pS470	Q9NZN4	EH-domain containing 2, isoform CRA_a	2.052	0.001
CRYBG1_pT397	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	2.032	0.009
FLNC_pS2146	Q14315	Filamin-C	1.893	0.029
LSM14A_pS192	Q8ND56	Isoform 2 of Protein LSM14 homolog A	1.839	0.164
NUP153_pS330	P49790	Nucleoporin 153kDa, isoform CRA_a	1.817	0.002
RAC1_pS71	P63000	Ras-related C3 botulinum toxin substrate 1 (Rho family, small GTP binding protein Rac1)	1.774	0.020
TNS4_pS253	Q8LZW8	Tensin-4	1.748	0.045
NA_pS177	B2RDZ9	cDNA, FLJ96850	1.738	0.011
PKP2_pS197	Q99959	Isoform 1 of Plakophilin-2	1.698	0.079
NUP153_pS333	P49790	Nucleoporin 153kDa, isoform CRA_a	1.696	0.002
CD2AP_pS458	Q9Y5K6	CD2-associated protein	1.694	0.001
E124_pS46	O14681	Etoposide induced 2.4 mRNA, isoform CRA_a	1.683	0.013
DSP_pS176	P15924	Desmoplakin	1.626	0.150
DSP_pS2815	P15924	Desmoplakin	1.601	0.014
DSP_pS2821	P15924	Desmoplakin	1.601	0.014
NA_pS60	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	1.476	0.008
NA_pS7	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	1.469	0.001
DSP_pS2825	P15924	Desmoplakin	1.467	0.099
HEL113_pS325	V9HWE1	Epididymis luminal protein 113	1.451	0.002
CRYBG1_pS424	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	1.426	0.001
CRYBG1_pS427	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	1.418	0.001
PAK2_pT143	Q13177	Serine/threonine-protein kinase PAK 2	1.394	0.004
SARG_pS133	Q9BW04	Specifically androgen-regulated gene protein	1.361	0.040
HEL-S-43_pS5	V9HW9	Protein S100	1.357	0.002
BIN1_pT292	O00499	Bridging integrator 1, isoform CRA_a	1.353	0.053
AKAP12_pS286	Q02952	A-kinase anchor protein 12	1.343	0.041
UBE1_pS46	P22314	Testicular secretory protein Li 63	1.334	0.001
NA_pS42	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	1.333	0.010
ITGB4_pS1069	P19144	Integrin beta	1.330	0.106
JUP_pS665	P14923	Junction plakoglobin isoform 1 (Fragment)	1.322	0.075
WDR44_pS561	Q5JSH3	Isoform 2 of WD repeat-containing protein 44	1.283	0.004
EPHA2_pS897	P29317	EPH receptor A2, isoform CRA_a	1.278	0.024
HIST1H4L_pS48	P62805	Histone H4	1.256	0.010
RAB7A_pS72	P51149	RAB7, member RAS oncogene family, isoform CRA_a	1.243	0.023
NA_pS266	B4E2X3	cDNA FLJ56024	1.240	0.007
RAB3GAP1_pS537	Q15042	RAB3 GTPase activating protein subunit 1	1.223	0.019
AKAP12_pS283	Q02952	A-kinase anchor protein 12	1.200	0.013
HACD3_pS135	Q9P035	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	1.197	0.005
BIN1_pS272	O00499	Bridging integrator 1, isoform CRA_a	1.189	0.004
PAK2_pS141	Q13177	Serine/threonine-protein kinase PAK 2	1.189	0.051
RPLP2_pS17	P05387	Ribosomal protein, large, P2, isoform CRA_a	1.162	0.011
HEL70_pS576	V9HW0C	Epididymis luminal protein 70	1.159	0.004
SNAP23_pS20	O00161	Synaptosomal-associated protein	1.156	0.009
EPHA2_pS901	P29317	EPH receptor A2, isoform CRA_a	1.149	0.012
PRKAR1A_pS83	P10644	Protein kinase, cAMP-dependent, regulatory, type I, alpha (Tissue specific extinguisher 1), isoform CRA_a	1.127	0.017
ARFGEF1_pS1569	Q9Y6D6	ADP-ribosylation factor guanine nucleotide-exchange factor 1 (Brefeldin A-inhibited), isoform CRA_a	1.122	0.012
HEL113_pS7	V9HWE1	Epididymis luminal protein 113	1.112	0.017
NA_pS1124	AOA1P7ZIM8	LMO7b	1.079	0.010
NA_pS1130	AOA1P7ZIM8	LMO7b	1.079	0.010
DDI2_pS194	Q5TDH0	Protein DDI1 homolog 2	1.074	0.017
TMEM201_pS454	Q5SNT2	Transmembrane protein 201	1.060	0.007
CTNND1_pS914	O60716	Isoform 1A of Catenin delta-1	1.059	0.011
ARHGAP35_pS1179	Q9NRY4	Rho GTPase-activating protein 35	1.051	0.033
DSG2_pS680	Q14126	Desmoglein-2	1.047	0.033
ZFYVE16_pS946	Q7Z3T8	Zinc finger, FYVE domain containing 16, isoform CRA_a	1.043	0.126
SCRIB_pS504	Q14160	Protein scribble homolog	1.026	0.032
PSAT1_pS331	Q9Y617	Phosphoserine aminotransferase	1.006	0.017



**Appendix Table 1 – continued.**

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2 (D492/D492M)	p Value
HEL-S-270_pS12	V9HW65	Annexin	-1.002	0.015
NA_pS2	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-1.004	0.023
PHIP_pS1315	Q8WWQ0	PH-interacting protein	-1.009	0.008
STOML2_pT327	Q9LJZ1	Stomatin-like protein 2, mitochondrial	-1.010	0.000
IGF2BP2_pS162	Q9Y6M1	Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 2	-1.011	0.002
GBF1_pS1780	Q92538	Isoform 3 of Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	-1.012	0.017
RPL4_pS295	P36578	60S ribosomal protein L4	-1.014	0.030
ADD3_pS649	Q9UEY8	Isoform 1 of Gamma-adducin	-1.018	0.048
UBE4B_pS105	Q95155	Isoform 2 of Ubiquitin conjugation factor E4 B	-1.018	0.045
SLK_pS777	Q9HG22	Isoform 2 of STE20-like serine/threonine-protein kinase	-1.020	0.048
NOP56_pS563	O00567	Nucleolar protein 56	-1.020	0.004
IGF2BP2_pS164	Q9Y6M1	Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 2	-1.025	0.001
LAMTOR1_pS27	Q6IAA8	Ragulator complex protein LAMTOR1	-1.026	0.024
EML3_pS176	Q32P44	Echinoderm microtubule-associated protein-like 3	-1.030	0.002
NUP50_pS193	Q9UKX7	Isoform 2 of Nuclear pore complex protein Nup50	-1.031	0.025
SMG8_pS742	Q8ND04	Protein SMG8	-1.033	0.019
BAG3_pS377	Q95817	BAG family molecular chaperone regulator 3	-1.037	0.006
SCRIB_pT1342	Q14160	Protein scribble homolog	-1.038	0.000
MAP4_pT1462	P27816	Microtubule-associated protein	-1.040	0.034
NA_pS139	A8K7N0	cDNA FLJ75556, highly similar to Homo sapiens ribosomal protein L14, mRNA	-1.041	0.021
YRDC_pS60	Q96U90	YrdC domain-containing protein, mitochondrial	-1.046	0.028
NA_pS413	B3KMW2	cDNA FLJ12778 fis, clone NT2RP2001740, moderately similar to Ubiquitin carboxyl-terminal hydrolase 36 (EC 3.1.2.15)	-1.047	0.033
DCP1A_pS487	Q9NPI6	Isoform 2 of mRNA-decapping enzyme 1A	-1.056	0.001
PKP3_pT571	Q9Y446	Plakophilin-3	-1.062	0.075
NA_pS204	B7Z3E3	Reticulon	-1.067	0.001
ARHGEF10_pS27	O15013	Rho guanine nucleotide exchange factor 10 (Fragment)	-1.067	0.084
SLK_pS779	Q9HG22	Isoform 2 of STE20-like serine/threonine-protein kinase	-1.081	0.048
RASAL2_pS877	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	-1.082	0.123
RASAL2_pS880	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	-1.082	0.123
PARG_pS178	Q86V56	Isoform 3 of Poly(ADP-ribose) glycohydrolase	-1.084	0.059
AHNAK_pS5752	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.086	0.057
UTP14A_pS437	Q9BVJ6	U3 small nucleolar RNA-associated protein 14 homolog A	-1.088	0.062
LIMD1_pS277	Q9UGP4	LIM domain-containing protein 1	-1.088	0.012
HTT_pS1874	P42858	Huntingtin	-1.096	0.019
ZNF106_pS1025	Q9H2Y7	Zinc finger protein 106	-1.101	0.111
ZNF106_pS1026	Q9H2Y7	Zinc finger protein 106	-1.101	0.111
IRS2_pT527	Q9Y4H2	Insulin receptor substrate 2	-1.109	0.050
IRS2_pS523	Q9Y4H2	Insulin receptor substrate 2	-1.109	0.050
PKP3_pS238	Q9Y446	Plakophilin-3	-1.123	0.014
TNKS1BP1_pS836	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.124	0.017
NA_pS14	A8K0D2	cDNA FLJ77740, highly similar to Homo sapiens 7-dehydrocholesterol reductase, mRNA	-1.126	0.005
PAK2_pS197	Q13177	Serine/threonine-protein kinase PAK 2	-1.127	0.023
NA_pS109	A0A1U9X609	ABCF1	-1.128	0.004
MAP7D1_pS125	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	-1.130	0.038
ERBB2_pS1024	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-1.131	0.045
SPAG9_pS265	O60271	C-Jun-amino-terminal kinase-interacting protein 4	-1.135	0.009
SPAG9_pS268	O60271	C-Jun-amino-terminal kinase-interacting protein 4	-1.135	0.009
SARM1_pS40	Q6SZW1	SARM1 protein (Fragment)	-1.137	0.052
EIF3E_pS399	P60228	Eukaryotic translation initiation factor 3 subunit E	-1.138	0.023
VCL_pS721	P18206	Vinculin, isoform CRA_c	-1.141	0.027
STIM2_pS680	Q9P246	Stromal interaction molecule 2	-1.142	0.043
STIM2_pS719	Q9P246	Stromal interaction molecule 2	-1.144	0.043
NA_pS56	A0A109NGN6	Proteasome subunit alpha type	-1.162	0.005
ZNF22_pS49	P17026	Zinc finger protein 22 (KOX 15), isoform CRA_a	-1.164	0.000
PKP3_pT308	Q9Y446	Plakophilin-3	-1.171	0.015
THRAP3_pS575	Q9Y2W1	Thyroid hormone receptor-associated protein 3	-1.176	0.018
LARP7_pS300	Q4G0J3	La-related protein 7	-1.184	0.006
NA_pS293	A8K2W3	cDNA FLJ78516	-1.185	0.024
TRAM1_pS279	Q15629	Translocating chain-associated membrane protein 1	-1.205	0.036
SH3KBP1_pS474	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-1.210	0.133
SH3KBP1_pS472	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-1.210	0.133
FKBP15_pS1164	Q5T1M5	FK506-binding protein 15	-1.214	0.000
MAP7D1_pS116	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	-1.228	0.001
ITGB4_pS1387	P16144	Integrin beta	-1.242	0.068
LARP1_pT1071	Q6PKG0	La-related protein 1	-1.246	0.015
MAP7D1_pS113	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	-1.262	0.138
AHNAK_pS3182	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.263	0.009



**Appendix Table 1 – continued.**

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2 (D492/D492M)	p Value
COBL1_pS256	Q53SF7	Isoform 4 of Cordon-bleu protein-like 1	-1.269	0.006
COBL1_pT260	Q53SF7	Isoform 4 of Cordon-bleu protein-like 1	-1.269	0.006
LARP7_pS299	Q4G0J3	La-related protein 7	-1.271	0.001
JUND_pS90	P17535	Transcription factor jun-D	-1.285	0.088
PEA15_pS116	Q15121	PEA15 protein	-1.292	0.002
MEF2D_pS121	Q14814	Isoform MEF2DA0 of Myocyte-specific enhancer factor 2D	-1.311	0.019
HSP90AB1_pS452	P08238	Heat shock protein 90kDa alpha (Cytosolic), class B member 1, isoform CRA_a	-1.317	0.041
OSBP_pS240	P22059	Oxysterol-binding protein 1	-1.328	0.000
NA_pT108	A0A1U9X609	ABCF1	-1.337	0.011
LARP7_pS298	Q4G0J3	La-related protein 7	-1.340	0.003
LIMA1_pS132	Q9UHB6	Isoform 4 of LIM domain and actin-binding protein 1	-1.347	0.002
VCL_pS290	P18206	Vinculin, isoform CRA_c	-1.362	0.014
KIF14_pS1292	Q15058	Kinesin-like protein KIF14	-1.362	0.034
NA_pS1166	Q2TTR7	Receptor protein-tyrosine kinase	-1.370	0.016
RPL28_pS115	P46779	60S ribosomal protein L28	-1.381	0.001
OSBPL3_pS251	Q9H4L5	Oxysterol-binding protein-related protein 3	-1.383	0.000
TEX2_pS196	Q81WB9	Testis-expressed protein 2	-1.396	0.120
ARHGAP1_pS51	Q07960	Rho GTPase-activating protein 1	-1.406	0.004
CCDC6_pS367	Q16204	Coiled-coil domain containing 6, isoform CRA_a	-1.406	0.009
PDHA1_pS293	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-1.431	0.003
NRBP1_pS11	Q9UHY1	Nuclear receptor-binding protein	-1.447	0.047
NRBP1_pS2	Q9UHY1	Nuclear receptor-binding protein	-1.447	0.053
DLG5_pT874	Q8TDM6	Isoform 4 of Disks large homolog 5	-1.458	0.002
DLG5_pS890	Q8TDM6	Isoform 4 of Disks large homolog 5	-1.458	0.002
DSG2_pS782	Q14126	Desmoglein-2	-1.476	0.023
RPS6KA4_pS347	O75676	Ribosomal protein S6 kinase	-1.484	0.002
RPS6KA4_pS343	O75676	Ribosomal protein S6 kinase	-1.484	0.002
SPECC1L_pS384	Q69YQ0	Isoform 2 of Cytospin-A	-1.499	0.019
ERRF1_pT127	Q9UJM3	ERBB receptor feedback inhibitor 1	-1.512	0.120
MYO18A_pS1970	Q92614	Unconventional myosin-XVIIIa	-1.519	0.003
NA_pS295	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-1.529	0.005
NA_pS292	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-1.529	0.194
AKAP13_pS1914	Q12802	Isoform 3 of A-kinase anchor protein 13	-1.564	0.019
SYNE1_pS7900	Q8NF91	Nesprin-1	-1.565	0.001
NA_pS16	A0A109NGN6	Proteasome subunit alpha type	-1.587	0.008
SPECC1L_pS385	Q69YQ0	Isoform 2 of Cytospin-A	-1.592	0.001
AHNAK_pS5400	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.612	0.120
EPHA2_pS570	P29317	EPH receptor A2, isoform CRA_a	-1.613	0.037
PDHA1_pS232	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-1.651	0.131
ERCC6L_pS1069	Q2NKX8	DNA excision repair protein ERCC-6-like	-1.684	0.153
HEL-S-102_pS15	V9HW43	Epididymis secretory protein Li 102	-1.688	0.024
GIT1_pS508	Q9Y2X7	ARF GTPase-activating protein GIT1	-1.734	0.010
PAWR_pS259	Q961Z0	PRKC apoptosis WT1 regulator protein	-1.748	0.042
NA_pS298	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-1.765	0.065
ERBB2_pS1077	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-1.806	0.001
AHNAK_pS5731	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.843	0.058
SPECC1L_pS389	Q69YQ0	Isoform 2 of Cytospin-A	-1.852	0.012
RAB23_pS188	Q9ULC3	RAB23, member RAS oncogene family, isoform CRA_a	-1.869	0.060
NA_pS148	Q71U35	Transcriptional enhancer factor TEF-5	-1.892	0.004
PKP2_pS155	Q99959	Isoform 1 of Plakophilin-2	-1.911	0.108
PKP2_pS154	Q99959	Isoform 1 of Plakophilin-2	-1.911	0.108
PKP2_pS151	Q99959	Isoform 1 of Plakophilin-2	-1.911	0.108
AHNAK_pT4766	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.919	0.037
MYO18A_pS1974	Q92614	Unconventional myosin-XVIIIa	-1.945	0.029
AHNAK_pT4100	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.953	0.017
ITGB4_pT1417	P16144	Integrin beta	-1.983	0.186
NA_pS268	B2R6N9	cDNA, FLJ93042, highly similar to Homo sapiens signal sequence receptor, alpha (translocon-associated protein alpha) (SSR1), mRNA	-1.983	0.002
ERRF1_pT131	Q9UJM3	ERBB receptor feedback inhibitor 1	-1.994	0.054
CAMK2D_pS319	Q13557	Isoform Delta 6 of Calcium/calmodulin-dependent protein kinase type II subunit delta	-2.007	0.028
CAMK2D_pS315	Q13557	Isoform Delta 6 of Calcium/calmodulin-dependent protein kinase type II subunit delta	-2.007	0.028
RPS6KA4_pS627	O75676	Ribosomal protein S6 kinase	-2.043	0.010
AHNAK_pS5857	Q09666	Neuroblast differentiation-associated protein AHNAK	-2.153	0.096
AHNAK_pT4430	Q09666	Neuroblast differentiation-associated protein AHNAK	-2.158	0.034
NOP56_pS569	O00567	Nucleolar protein 56	-2.208	0.000
NOP56_pS570	O00567	Nucleolar protein 56	-2.208	0.000
AHNAK_pS5830	Q09666	Neuroblast differentiation-associated protein AHNAK	-2.213	0.092
CTNND1_pS47	O60716	Isoform 1A of Catenin delta-1	-2.263	0.005
AHNAK_pT3716	Q09666	Neuroblast differentiation-associated protein AHNAK	-2.277	0.040
RRAGC_pT96	Q9HB90	Ras-related GTP-binding protein C	-2.387	0.001
CTNND1_pS191	P35222	Catenin (Cadherin-associated protein), beta 1, 88kDa, isoform CRA_a	-2.393	0.002
ZNF106_pS1370	Q9HZY7	Zinc finger protein 106	-2.409	0.004
PDHA1_pS300	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-2.413	0.058
SH3KBP1_pS193	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-2.449	0.012
CTNND1_pS4	O60716	Isoform 1A of Catenin delta-1	-2.486	0.014
ITGB4_pS1404	P16144	Integrin beta	-2.545	0.087
AHNAK_pS5620	Q09666	Neuroblast differentiation-associated protein AHNAK	-2.627	0.008
CDH1_pS793	P12830	E-cadherin 1	-2.631	0.016
NA_pT290	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-2.660	0.001
CAMSAP2_pS835	O08AD1	Isoform 2 of Calmodulin-regulated spectrin-associated protein 2	-3.479	0.018
ABLIM3_pS282	O94929	Isoform 2 of Actin-binding LIM protein 3	-3.567	0.013
ITGB4_pS1534	P16144	Integrin beta	-4.577	0.003



**Appendix Table 2.** The topmost dysregulated phosphorylation sites by comparing D492HER2 with D492 after statistical analysis of the SILAC phosphoproteomic dataset (fold change  $\geq 2$ , one-sample Student's T-test p value  $< 0.05$ ;  $p > 0.05$  and fold change  $\geq 2$  in all three replicates). Tenne: active in D492HER2; Blue: active in D492. NA: gene name is not available.

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492 HER2/D492)	p Value
RPS6KA4_pS627	O75676	Ribosomal protein S6 kinase	5.198	0.005
CAV1_pY6	Q03135	Caveolin	2.862	0.028
AHNAK_pS5731	Q09666	Neuroblast differentiation-associated protein AHNAK	2.702	0.160
RASAL2_pS880	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	2.679	0.073
RASAL2_pS877	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	2.679	0.073
AHNAK_pS5830	Q09666	Neuroblast differentiation-associated protein AHNAK	2.646	0.088
RRAGC_pT96	Q9HB90	Ras-related GTP-binding protein C	2.643	0.001
AHNAK_pS5857	Q09666	Neuroblast differentiation-associated protein AHNAK	2.173	0.082
ERBB2_pS968	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	1.947	0.011
CDC42EP2_pS109	O14613	Cdc42 effector protein 2	1.946	0.003
HEL-S-270_pS12	V9HW65	Annexin	1.848	0.002
CCDC6_pS367	Q16204	Coiled-coil domain containing 6, isoform CRA_a	1.811	0.005
MYO18A_pS1974	Q92614	Unconventional myosin-XVIIIa	1.810	0.013
MYO18A_pS1970	Q92614	Unconventional myosin-XVIIIa	1.810	0.013
CDH1_pS793	P12830	E-cadherin 1	1.788	0.025
ITGB4_pS1534	P16144	Integrin beta	1.713	0.037
CTNNB1_pS191	P35222	Catenin (Cadherin-associated protein), beta 1, 88kDa, isoform CRA_a	1.643	0.019
AHNAK_pS5400	Q09666	Neuroblast differentiation-associated protein AHNAK	1.627	0.159
AKAP13_pS1914	Q12802	Isoform 3 of A-kinase anchor protein 13	1.491	0.009
MAP4_pT1462	P27816	Microtubule-associated protein	1.483	0.004
HEL-S-270_pY30	V9HW65	Annexin	1.416	0.001
CTNND1_pS4	O60716	Isoform 1A of Catenin delta-1	1.415	0.034
NA_pT290	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	1.410	0.091
MAP7D1_pS116	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	1.398	0.000
IRS2_pS388	Q9Y4H2	Insulin receptor substrate 2	1.373	0.012
COBLL1_pS256	Q53SF7	Isoform 4 of Cordon-bleu protein-like 1	1.370	0.089
COBLL1_pT260	Q53SF7	Isoform 4 of Cordon-bleu protein-like 1	1.370	0.089
MAP7D1_pS113	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	1.366	0.123
MAP7D1_pS125	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	1.366	0.073
ERBB2_pS1144	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	1.364	0.030
ADD3_pS645	Q9UEY8	Isoform 1 of Gamma-adducin	1.362	0.078
ADD3_pS649	Q9UEY8	Isoform 1 of Gamma-adducin	1.362	0.097
STOML2_pT327	Q9UJZ1	Stomatin-like protein 2, mitochondrial	1.343	0.006
TFAM_pS195	Q00059	TFAM protein (Fragment)	1.332	0.002
NOP56_pS569	O00567	Nucleolar protein 56	1.324	0.001
NOP56_pS570	O00567	Nucleolar protein 56	1.324	0.001
NOP56_pS563	O00567	Nucleolar protein 56	1.324	0.001
PAK2_pS197	Q13177	Serine/threonine-protein kinase PAK 2	1.317	0.029
OSBP_pS240	P22059	Oxysterol-binding protein 1	1.314	0.005
PLEKHA5_pS471	Q9HAAU	Pleckstrin homology domain-containing family A member 5	1.299	0.044
SF3B1_pT328	O75533	Splicing factor 3B subunit 1	1.297	0.069
DSG2_pS782	Q14126	Desmoglein-2	1.295	0.032
ERBIN_pS1158	Q96RT1	Isoform 5 of ErbB1	1.293	0.037
PAWR_pS259	Q96I20	PRKC apoptosis WT1 regulator protein	1.260	0.032
ERBB2_pT671	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	1.256	0.196
ERRF1_pT131	Q9UJM3	ERBB receptor feedback inhibitor 1	1.237	0.051
GIT1_pS508	Q9Y2X7	ARF GTPase-activating protein GIT1	1.235	0.043
NA_pS293	A8K2W3	cDNA FLJ78516	1.221	0.015
MAP4_pS723	P27816	Microtubule-associated protein	1.202	0.041
NRBP1_pS11	Q9UHY1	Nuclear receptor-binding protein	1.196	0.069
NRBP1_pS2	Q9UHY1	Nuclear receptor-binding protein	1.196	0.070
NA_pY1172	Q2TTR7	Receptor protein-tyrosine kinase	1.164	0.042
NOP56_pS519	O00567	Nucleolar protein 56	1.164	0.026
NOP56_pS520	O00567	Nucleolar protein 56	1.164	0.026
ERRF1_pT127	Q9UJM3	ERBB receptor feedback inhibitor 1	1.157	0.094
SLC8A1_pS389	P32418	Sodium/calcium exchanger 1	1.143	0.096
TNS4_pS7	Q8IZW8	Tensin-4	1.137	0.027
CTNND1_pS47	O60716	Isoform 1A of Catenin delta-1	1.120	0.005
PRRC2C_pT2682	Q9Y520	Isoform 4 of Protein PRRC2C	1.085	0.084
PPP1R13L_pS187	Q8WUF5	Protein phosphatase 1, regulatory (Inhibitor) subunit 13 like, isoform CRA_a	1.085	0.016
RAPH1_pS853	Q70E73	Ras-associated and pleckstrin homology domains-containing protein 1	1.083	0.011
NA_pT108	AA01U9X609	ABCF1	1.072	0.005
NA_pY1110	Q2TTR7	Receptor protein-tyrosine kinase	1.071	0.010
AHNAK_pS5448	Q09666	Neuroblast differentiation-associated protein AHNAK	1.063	0.107
DCP1A_pS487	Q9NPI6	Isoform 2 of mRNA-decapping enzyme 1A	1.062	0.029
SF3B2_pS307	Q13435	Splicing factor 3B subunit 2	1.054	0.048

## Appendix Table 2 – continued.

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492 HER2/D492)	p Value
HACD3_pS135	Q9P035	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	-1.003	0.023
AHNAK_pS210	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.006	0.154
TNS4_pS350	Q8IZW8	Tensin-4	-1.007	0.082
PKP2_pS155	Q99959	Isoform 1 of Plakophilin-2	-1.009	0.000
PKP2_pS151	Q99959	Isoform 1 of Plakophilin-2	-1.015	0.017
NUP153_pS333	P49790	Nucleoporin 153kDa, isoform CRA_a	-1.016	0.003
CTNND1_pS352	O60716	Isoform 1A of Catenin delta-1	-1.017	0.014
PANK2_pS45	Q9BZ23	Isoform 2 of Pantothenate kinase 2, mitochondrial	-1.039	0.018
FAM83H_pS513	Q6ZRV2	Protein FAM83H	-1.041	0.013
TACC3_pS25	Q9Y6A5	Transforming acidic coiled-coil-containing protein 3	-1.044	0.078
CTNND1_pS346	O60716	Isoform 1A of Catenin delta-1	-1.045	0.100
CTNND1_pS349	O60716	Isoform 1A of Catenin delta-1	-1.045	0.100
MAP7D1_pS428	Q3KQU3	Isoform 4 of MAP7 domain-containing protein 1	-1.049	0.009
TMEM201_pS454	Q5SNT2	Transmembrane protein 201	-1.058	0.069
ALDOC_pS45	P09972	Fructose-bisphosphate aldolase	-1.063	0.001
EPS8L2_pS17	Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2	-1.069	0.073
TMPO_pS362	P42166	Thymopoietin, isoform CRA_c	-1.071	0.004
GPS1_pS454	Q13098	COP9 signalosome complex subunit 1	-1.074	0.021
CTNND1_pS252	O60716	Isoform 1A of Catenin delta-1	-1.088	0.037
CRYBG1_pS427	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	-1.102	0.000
DSP_pS22	P15924	Desmoplakin	-1.104	0.098
TAKO1_pS9	Q7L7X3	TAO kinase 1, isoform CRA_a	-1.108	0.025
NA_pS350	B2RC16	cDNA, FLJ96094, highly similar to Homo sapiens numb homolog (Drosophila) (NUMB), mRNA	-1.108	0.112
STK38_pS264	Q15208	Serine/threonine kinase 38, isoform CRA_a	-1.112	0.189
UBE1_pS46	P22314	Testicular secretory protein Li 63	-1.115	0.002
NUP153_pS330	P49790	Nucleoporin 153kDa, isoform CRA_a	-1.120	0.021
EHD2_pS470	Q9NZN4	EH-domain containing 2, isoform CRA_a	-1.126	0.001
CRYBG1_pS424	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	-1.129	0.000
IRS2_pS1203	Q9Y4H2	Insulin receptor substrate 2	-1.133	0.012
NA_pS266	B4E2X3	cDNA FLJ56024	-1.139	0.002
GPS1_pS448	Q13098	COP9 signalosome complex subunit 1	-1.160	0.014
AKAP12_pS283	Q02952	A-kinase anchor protein 12	-1.165	0.015
IRS2_pS973	Q9Y4H2	Insulin receptor substrate 2	-1.165	0.126
ATXN2_pS545	Q99700	Ataxin-2 (Fragment)	-1.171	0.003
TNKS1BP1_pS987	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.175	0.005
HEL113_pS325	V9HWE1	Epididymis luminal protein 113	-1.186	0.005
NES_pS398	P48681	Nestin, isoform CRA_c	-1.190	0.170
RAC1_pS71	P63000	Ras-related C3 botulinum toxin substrate 1 (Rho family, small GTP binding protein Rac1)	-1.213	0.047
ZFYVE16_pS946	Q7Z3T8	Zinc finger, FYVE domain containing 16, isoform CRA_a	-1.217	0.033
FAM62A_pS820	Q9BSJ8	Family with sequence similarity 62 (C2 domain containing), member A, isoform CRA_b	-1.221	0.003
CHORDC1_pT47	Q9UHD1	Cysteine and histidine-rich domain (CHORD)-containing 1, isoform CRA_c	-1.233	0.014
PHKA2_pS729	P46019	Phosphorylase b kinase regulatory subunit alpha, liver isoform	-1.239	0.058
PSAT1_pS344	Q9Y617	Phosphoserine aminotransferase	-1.299	0.041
DSP_pS176	P15924	Desmoplakin	-1.308	0.002
CTNND1_pS268	O60716	Isoform 1A of Catenin delta-1	-1.309	0.022
CTNND1_pS269	O60716	Isoform 1A of Catenin delta-1	-1.309	0.022
HEL113_pS42	V9HWE1	Epididymis luminal protein 113	-1.370	0.007

## Appendix Table 2 – continued.

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492 HER2/D492)	p Value
WDR44_pS561	Q5JSH3	Isoform 2 of W/D repeat-containing protein 44	-1.410	0.004
PKP2_pS251	Q99959	Isoform 1 of Plakophilin-2	-1.427	0.200
CTNND1_pS230	O60716	Isoform 1A of Catenin delta-1	-1.438	0.023
HIST1H4L_pS48	P62805	Histone H4	-1.447	0.017
RPL34_pS12	P49207	Ribosomal protein L34, isoform CRA_a	-1.460	0.151
AKAP12_pS286	Q02952	A-kinase anchor protein 12	-1.469	0.034
TNS3_pS1149	Q68CZ2	Tensin-3	-1.487	0.044
TNKS1BP1_pS1024	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.493	0.033
ITPR3_pS916	Q14573	Inositol 1,4,5-trisphosphate receptor type 3	-1.503	0.013
HEL113_pS56	V9HWE1	Epididymis luminal protein 113	-1.516	0.009
HEL113_pS51	V9HWE1	Epididymis luminal protein 113	-1.516	0.009
BIN1_pT292	O00499	Bridging integrator 1, isoform CRA_a	-1.522	0.118
AHNAK_pS332	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.548	0.012
NA_pS177	B2RDZ9	cDNA, FLJ96850	-1.566	0.000
SEPT9_pS12	A0A024R8V0	Septin 9, isoform CRA_a	-1.571	0.059
DST_pS1909	Q03001	Isoform 3 of Dystonin	-1.587	0.046
EI24_pS46	O14681	Etoposide induced 2.4 mRNA, isoform CRA_a	-1.625	0.010
SCRIB_pS504	Q14160	Protein scribble homolog	-1.636	0.010
NA_pS60	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	-1.669	0.003
ITGB4_pT1417	P16144	Integrin beta	-1.679	0.109
SARG_pS133	Q9BW04	Specifically androgen-regulated gene protein	-1.774	0.156
AKAP12_pS75	Q02952	A-kinase anchor protein 12	-1.797	0.173
FLNC_pS2146	Q14315	Filamin-C	-1.889	0.007
NA_pS42	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	-1.918	0.002
PKP2_pS197	Q99959	Isoform 1 of Plakophilin-2	-1.920	0.001
LMNB2_pS424	Q03252	Lamin-B2	-1.999	0.001
COL17A1_pS148	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	-2.041	0.009
ITGB4_pS1413	P16144	Integrin beta	-2.139	0.015
NA_pS10	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	-2.296	0.024
ITGB4_pS1069	P16144	Integrin beta	-2.303	0.009
NA_pS31	B2RA03	cDNA, FLJ94640, highly similar to Homo sapiens keratin 18 (KRT18), mRNA	-2.304	0.011
CRYBG1_pS280	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	-2.424	0.029
AKAP12_pS96	Q02952	A-kinase anchor protein 12	-2.479	0.069
FAM83B_pS869	Q5T0W9	Protein FAM83B	-2.516	0.021
CRYBG1_pT397	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	-2.563	0.003
AKAP12_pS1395	Q02952	A-kinase anchor protein 12	-2.848	0.001
ERBB2_pY975	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-2.880	0.011
CRYBG1_pS299	Q9Y4K1	Beta/gamma crystallin domain-containing protein 1	-2.964	0.000
ERBB2_pS1070	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.622	0.011
ERBB2_pS1024	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.660	0.001
ITGB4_pS1424	P16144	Integrin beta	-3.760	0.000
ERBB2_pS1036	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.815	0.006
ERBB2_pS1048	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-4.040	0.012



**Appendix Table 3.** The topmost dysregulated phosphorylation sites by comparing D492HER2 with D492 after statistical analysis of the SILAC phosphoproteomic dataset (fold change  $\geq 2$ , one-sample Student's T-test p value  $< 0.05$ ;  $p > 0.05$  and fold change  $\geq 2$  in all three replicates). Tenne: active in D492HER2; Red: active in D492M. NA: gene name is not available.

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492 HER2/D492M)	p Value
COL17A1_pS93	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	3.886	0.086
CAV1_pY6	Q03135	Caveolin	3.692	0.013
COL17A1_pS148	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	3.467	0.017
COL17A1_pS61	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	3.295	0.146
RPS6KA4_pS627	O75676	Ribosomal protein S6 kinase	3.088	0.037
COL17A1_pS85	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	2.904	0.012
CD2AP_pS458	Q9Y5K6	CD2-associated protein	2.626	0.117
COL17A1_pS62	Q9UMD9	Isoform 2 of Collagen alpha-1(XVII) chain	2.597	0.025
TNS4_pS253	Q8IZW8	Tensin-4	2.257	0.083
NA_pS1124	A0A1P7ZIM8	LMO7b	1.715	0.014
NA_pS1130	A0A1P7ZIM8	LMO7b	1.715	0.014
LRBA_pS2485	P50851	Isoform 2 of Lipopolysaccharide-responsive and beige-like anchor protein	1.660	0.030
CTNND1_pS320	O60716	Isoform 1A of Catenin delta-1	1.590	0.074
DSP_pS2815	P15924	Desmoplakin	1.516	0.033
DSP_pS2821	P15924	Desmoplakin	1.516	0.033
DSP_pS2825	P15924	Desmoplakin	1.516	0.097
RAB7A_pS72	P51149	RAB7, member RAS oncogene family, isoform CRA_a	1.508	0.064
EPHA2_pS901	P29317	EPH receptor A2, isoform CRA_a	1.498	0.008
CEP170B_pS1109	Q9Y4F5	Isoform 3 of Centrosomal protein of 170 kDa protein B	1.392	0.008
PKP3_pS313	Q9Y446	Plakophilin-3	1.349	0.093
RASAL2_pS880	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	1.334	0.059
RASAL2_pS877	Q9UJF2	Isoform 2 of Ras GTPase-activating protein nGAP	1.334	0.059
SNAP23_pS20	O00161	Synaptosomal-associated protein	1.246	0.021
ERBB2_pS968	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	1.179	0.073
TNS4_pS350	Q8IZW8	Tensin-4	1.164	0.188
ZC3HAV1_pS335	Q7Z2W4	Zinc finger CCCH-type antiviral protein 1	1.160	0.018
EPHA2_pS897	P29317	EPH receptor A2, isoform CRA_a	1.122	0.006
PRKAR1A_pS83	P10644	Protein kinase, cAMP-dependent, regulatory, type I, alpha (Tissue specific extinguisher 1), isoform CRA_a	1.106	0.016
PANK4_pT406	Q9NVE7	Pantothenate kinase 4	1.105	0.064
CHD4_pS1524	Q14839	Chromodomain-helicase-DNA-binding protein 4	1.067	0.096
CHD4_pS1528	Q14839	Chromodomain-helicase-DNA-binding protein 4	1.067	0.096
EPHA2_pS892	P29317	EPH receptor A2, isoform CRA_a	1.041	0.119
SETX_pS1019	Q7Z333	Isoform 3 of Probable helicase senataxin	1.033	0.148
FAM62A_pS820	Q9BSJ8	Family with sequence similarity 62 (C2 domain containing), member A, isoform CRA_b	1.029	0.000
JAM3_pS230	Q9BX67	Isoform 2 of Junctional adhesion molecule C	1.028	0.003
NA_pY1172	Q2TTR7	Receptor protein-tyrosine kinase	1.020	0.097
NA_pS73	B3KN59	cDNA FLJ13673 fis, clone PLACE1011858, highly similar to BAG family molecular chaperone regulator 2	1.013	0.004
HEL-S-270_pY30	V9HW65	Annexin	1.013	0.013

**Appendix Table 3 – continued.**

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492 HER2/D492M)	p Value
CTNND1_pT900	O60716	Isoform 1A of Catenin delta-1	-1.004	0.015
TNKS1BP1_pT1282	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.005	0.070
HTT_pS432	P42858	Huntingtin	-1.005	0.022
DKFZp76110921_pS692	Q68CX0	Transporter (Fragment)	-1.006	0.099
PPP1R13L_pS102	Q8WUF5	Protein phosphatase 1, regulatory (Inhibitor) subunit 13 like, isoform CRA_a	-1.011	0.003
CDAN1_pS276	Q8IWY9	Isoform 1 of Codanin-1	-1.026	0.055
NA_pS298	B3KU62	cDNA FLJ39243 fis, clone OCBBF2008283, highly similar to Protein NDRG1	-1.026	0.116
TOP2A_pS1374	P11388	DNA topoisomerase 2-alpha	-1.031	0.139
TOP2A_pS1377	P11388	DNA topoisomerase 2-alpha	-1.031	0.139
BAIAP2L1_pS261	Q9UHR4	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	-1.031	0.001
MST065_pS15	Q549C5	HCG2010808, isoform CRA_a	-1.036	0.007
SAP30BP_pS22	Q9UHR5	Isoform 2 of SAP30-binding protein	-1.040	0.055
MYBBP1A_pS1308	Q9BQG0	Myb-binding protein 1A	-1.040	0.004
MYBBP1A_pS1314	Q9BQG0	Myb-binding protein 1A	-1.040	0.004
STIM2_pS680	Q9P246	Stromal interaction molecule 2	-1.040	0.003
PDLIM2_pS161	Q96JV6	PDZ and LIM domain protein 2	-1.042	0.014
TNKS1BP1_pS1024	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.043	0.044
CAD_pS1343	P27708	CAD protein	-1.048	0.102
PLEKHF2_pS16	Q9H8V4	Pleckstrin homology domain-containing family F member 2	-1.049	0.006
SLK_pS779	Q9H2G2	Isoform 2 of STE20-like serine/threonine-protein kinase	-1.051	0.070
SLK_pS777	Q9H2G2	Isoform 2 of STE20-like serine/threonine-protein kinase	-1.051	0.054
NAP1L4_pS5	Q99733	Nucleosome assembly protein 1-like 4, isoform CRA_b	-1.053	0.021
HEL113_pS42	V9HWE1	Epididymis luminal protein 113	-1.054	0.022
TNKS1BP1_pS1029	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.068	0.046
TACC3_pS25	Q9Y6A5	Transforming acidic coiled-coil-containing protein 3	-1.076	0.003
NOP56_pS563	O00567	Nucleolar protein 56	-1.078	0.019
SH3KBP1_pS474	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-1.080	0.056
SH3KBP1_pS472	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-1.080	0.056
SARM1_pS40	Q6SZW1	SARM1 protein (Fragment)	-1.084	0.043
REPS1_pS272	Q96D71	Isoform 3 of RaiBP1-associated Eps domain-containing protein 1	-1.092	0.077
REPS1_pS273	Q96D71	Isoform 3 of RaiBP1-associated Eps domain-containing protein 1	-1.092	0.077
HEL-S-49_pS21	V9HWK1	Triosephosphate isomerase	-1.094	0.003
LASP1_pS146	Q14847	LIM and SH3 protein 1, isoform CRA_b	-1.097	0.128
TP53BP1_pS1618	Q12888	TP53-binding protein 1	-1.106	0.008
YRDC_pS60	Q86U90	YrdC domain-containing protein, mitochondrial	-1.106	0.014
EIF4B_pS498	P23588	Eukaryotic translation initiation factor 4B	-1.110	0.004
LASP1_pT68	Q14847	LIM and SH3 protein 1, isoform CRA_b	-1.113	0.021
SPECC1L_pS384	Q69YQ0	Isoform 2 of Cytospin-A	-1.117	0.040
PKP3_pT308	Q9Y446	Plakophilin-3	-1.123	0.143
IRS2_pS560	Q9Y4H2	Insulin receptor substrate 2	-1.127	0.040
IGF2BP2_pS164	Q9Y6M1	Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 2	-1.129	0.001
CTNND1_pS352	O60716	Isoform 1A of Catenin delta-1	-1.137	0.000
CTNND1_pS349	O60716	Isoform 1A of Catenin delta-1	-1.137	0.179
AHNAK_pS210	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.144	0.108
TNKS1BP1_pS987	Q9C0C2	Tankyrase 1 binding protein 1, 182kDa, isoform CRA_a	-1.147	0.050
C15orf52_pS133	Q6ZU76	Isoform 4 of Uncharacterized protein C15orf52	-1.169	0.187
ITPR3_pS916	Q14573	Inositol 1,4,5-trisphosphate receptor type 3	-1.179	0.071
PDLIM4_pS120	P50479	PDZ and LIM domain protein 4	-1.183	0.006
NA_pS185	B3KN57	cDNA FLJ13654 fis, clone PLACE1011477, highly similar to Sorting nexin-2	-1.192	0.017
IGF2BP2_pS162	Q9Y6M1	Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 2	-1.193	0.003



**Appendix Table 3 – continued.**

Phosphorylation Site	Uniprot Protein ID	Protein Name	Median_Log2(D492/HER2/D492M)	p Value
AHNAK_pS552	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.201	0.016
TMPO_pS362	P42166	Thymopoietin, isoform CRA_c	-1.203	0.009
CENPF_pS2996	P49454	Centromere protein F	-1.204	0.011
CTNND1_pS230	O60716	Isoform 1A of Catenin delta-1	-1.224	0.030
NA_pS1166	Q2TTR7	Receptor protein-tyrosine kinase	-1.225	0.079
HEL113_pS56	V9HWE1	Epididymis luminal protein 113	-1.230	0.034
HEL113_pS51	V9HWE1	Epididymis luminal protein 113	-1.230	0.034
NA_pS14	A8K0D2	cDNA FLJ77740, highly similar to Homo sapiens 7-dehydrocholesterol reductase, mRNA	-1.242	0.003
SBDS_pT5	Q9Y3A5	Shwachman-Bodian-Diamond syndrome isoform 1 (Fragment)	-1.245	0.007
STIM2_pS719	Q9P246	Stromal interaction molecule 2	-1.250	0.004
NAP1L4_pS7	Q99733	Nucleosome assembly protein 1-like 4, isoform CRA_b	-1.251	0.018
HNRNPA3_pS14	P51991	Heterogeneous nuclear ribonucleoprotein A3	-1.258	0.066
CTNND1_pS47	O60716	Isoform 1A of Catenin delta-1	-1.261	0.014
PEA15_pS116	Q15121	PEA15 protein	-1.262	0.005
NAP1L4_pS12	Q99733	Nucleosome assembly protein 1-like 4, isoform CRA_b	-1.274	0.001
CENPF_pS3007	P49454	Centromere protein F	-1.334	0.017
PDHA1_pS293	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-1.347	0.006
PHACTR4_pS411	Q8I221	Isoform 3 of Phosphatase and actin regulator 4	-1.359	0.054
NA_pS16	A0A109NGN6	Proteasome subunit alpha type	-1.387	0.004
ERBB2_pS1053	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-1.390	0.056
RIPOR1_pS22	Q6ZS17	Isoform 2 of Rho family-interacting cell polarization regulator 1	-1.391	0.001
VCL_pS721	P18206	Vinculin, isoform CRA_c	-1.400	0.000
HSP90AB1_pS452	P08238	Heat shock protein 90kDa alpha (Cytosolic), class B member 1, isoform CRA_a	-1.418	0.027
TEX2_pS196	Q8IWB9	Testis-expressed protein 2	-1.421	0.002
PKP3_pS314	Q9Y446	Plakophilin-3	-1.421	0.102
AHNAK_pS5620	Q09666	Neuroblast differentiation-associated protein AHNAK	-1.426	0.107
HEL-S-102_pS15	V9HW43	Epididymis secretory protein Li 102	-1.443	0.149
SPECC1L_pS385	Q69YQ0	Isoform 2 of Cytospin-A	-1.521	0.014
PDLIM2_pS134	Q96JY6	PDZ and LIM domain protein 2	-1.535	0.006
PDLIM2_pS137	Q96JY6	PDZ and LIM domain protein 2	-1.535	0.015
TP53BP1_pS1678	Q12888	TP53-binding protein 1	-1.544	0.016
TP53BP1_pS1673	Q12888	TP53-binding protein 1	-1.544	0.016
NA_pS148	Q71U35	Transcriptional enhancer factor TEF-5	-1.565	0.002
PDHA1_pS232	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-1.612	0.030
TNS3_pS1154	Q68CZ2	Tensin-3	-1.623	0.019
TNS3_pS1149	Q68CZ2	Tensin-3	-1.623	0.019
PKP3_pT571	Q9Y446	Plakophilin-3	-1.625	0.028
SPECC1L_pS389	Q69YQ0	Isoform 2 of Cytospin-A	-1.700	0.008
EIF3E_pS399	P60228	Eukaryotic translation initiation factor 3 subunit E	-1.750	0.001
ITGB4_pS1413	P16144	Integrin beta	-1.845	0.125
PLCB3_pS537	Q01970	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3	-1.887	0.125
ERCC6L_pS1069	Q2NKX8	DNA excision repair protein ERCC-6-like	-1.919	0.199
ZNF106_pS1370	Q9H2Y7	Zinc finger protein 106	-1.951	0.001
PDHA1_pS300	P08559	Pyruvate dehydrogenase E1 component subunit alpha	-1.980	0.134
SEPT9_pS12	A0A024R8V0	Septin 9, isoform CRA_a	-1.990	0.014
PKP2_pS251	Q99959	Isoform 1 of Plakophilin-2	-2.006	0.149
VCL_pS290	P18206	Vinculin, isoform CRA_c	-2.048	0.002
RAB23_pS188	Q9ULC3	RAB23, member RAS oncogene family, isoform CRA_a	-2.352	0.001
SH3KBP1_pS193	Q96B97	Isoform 2 of SH3 domain-containing kinase-binding protein 1	-2.548	0.008
ITGB4_pS1424	P16144	Integrin beta	-3.174	0.092
ERBB2_pS1077	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.303	0.005
CAMSAP2_pS835	Q08AD1	Isoform 2 of Calmodulin-regulated spectrin-associated protein 2	-3.522	0.006
ABLIM3_pS282	O94929	Isoform 2 of Actin-binding LIM protein 3	-3.525	0.016
ERBB2_pY975	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.544	0.001
ERBB2_pY847	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-3.734	0.019
ERBB2_pS1048	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-4.226	0.005
ERBB2_pS1070	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-4.897	0.004
ERBB2_pS1024	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-5.930	0.006
ERBB2_pS1036	P04626	Isoform 5 of Receptor tyrosine-protein kinase erbB-2	-6.706	0.029