Metabolic systems analysis of breast epithelial cells following epithelial-to-mesenchymal transition

Sigurður Trausti Karvelsson

Thesis for the degree of Philosophiae Doctor

Supervisors:
Dr. Óttar Rolfsson

Dr. Steinn Guðmundsson

Doctoral committee:
Dr. Skarphéðinn Halldórsson

Dr. Margrét Þorsteinsdóttir

Dr. Þórunn Rafnar

February 2022

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

Kerfislíffræðileg efnaskiptagreining á brjóstaþekjufrumum við bandvefsumbreytingu

Sigurður Trausti Karvelsson

Ritgerð til doktorsgráðu

Leiðbeinendur:
Dr. Óttar Rolfsson

Dr. Steinn Guðmundsson

Doktorsnefnd:
Dr. Skarphéðinn Halldórsson

Dr. Margrét Þorsteinsdóttir

Dr. Þórunn Rafnar

Febrúar 2022

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

ISBN 978-9935-9657-0-7

Printing by Háskólaprent

Reykjavik, Iceland 2022

Ágrip

Brjóstakrabbamein er margþættur sjúkdómur þar sem boðskiptaferli, genatjáning, efnaskiptaferli og vaxtarferli í brjóstavef eru gölluð. Sjúkdómurinn veldur dauða yfir 600,000 manns í heiminum á ári hverju. Með því að rannsaka líffræðileg ferli í eðlilegum brjóstavef fæst betri innsýn í meinafræði brjóstakrabbameins. Í þessari ritgerð var einblínt á þroskunarfræðilegt ferli sem nefnist bandvefsumbreyting sem talin er nauðsynlegt fyrir myndun meinvarpa. Sérstaklega var einblínt á efnaskiptabreytingar samhliða þessu ferli í brjóstaþekjufrumum. Ritgerðin skiptist í þrjá hluta. Í fyrsta hlutanum var ^{13}C samsætugreining notuð til að skilgreina efnaskiptabreytingar sem gerast samhliða bandvefsumbreytingu D492 brjóstaþekjufrumulínunnar. Niðurstöður sýndu að frumulínan minnkaði sykurrofsvirkni og jók afoxandi karboxýleringu amínósýrunnar glútamíns. Aukinheldur breyttust efnaskipti oxunar/afoxunarhvarfa, þar sem meðal annars myndun glútaþíóns minnkaði umtalsvert sem hafði bein áhrif á lyfjanæmni frumnanna. Í öðrum hluta ritgerðarinnar var notkunargildi efnaskiptalíkana til að spá fyrir um efnaskipti D492 frumulínunnar fyrir og eftir bandvefsumbreytingu skoðuð. Mismunandi mengjagögn voru notuð til að skorða líkönin. Líkön skorðuð með próteinmengjagögnum reyndust nákvæmust í spám á flæðigildum efnahvarfa, samanborið við niðurstöður úr ^{13}C samsætugreiningum. Frekari greining á líkönum skorðuðum með próteinmengjagögnum sýndi fram á nýtanlega veiklun í efnaskiptum frumnanna sem staðfest var með frumulíffræðitilraunum. Að lokum voru próteinmengjagögn úr æxlum brjóstakrabbameinssjúklinga notuð til að skorða efnaskiptalíkanið. Frekari greining á þeim líkönum sýndi fram á mögulega lífvísa fyrir lifun brjóstakrabbameinssjúklinga í mismunandi undirflokkum. Í þriðja og síðasta hluta ritgerðarinnar var hlutverk GFPT2 ensímsins í bandvefsumbreytingu og æxlismyndun í brjóstavef rannsakað. Greining á próteinmengi D492 frumulínunnar (i) fyrir bandvefsumbreytingu, (ii) eftir bandvefsumbreytingu og (iii) eftir oftjáningu æxlisgensins HER2 leiddi í ljós jákvætt samband milli próteinmagns GFPT2 ensímsins og bandvefsumbreytingarstöðu frumulínanna. Frekari sameindalíffræðilegar og lífupplýsingafræðilegar rannsóknir bentu á að GFPT2 væri mikilvægt fyrir fjölsykruefnaskipti og svörun frumnanna við oxunarálagi við bandvefsumbreytingu og æxlismyndun.

Niðurstöður verkefnisins hafa aukið skilning okkar á þeim efnaskiptabreytingum sem eiga sér stað við bandvefsumbreytingu brjóstaþekjufrumna og hvernig þær hafa áhrif á lyfjanæmni þeirra. Einnig er hér sýnt fram á hve gagnsöm efnaskiptalíkön skorðuð af mengjagögnum geta verið þegar kemur að því að rannsaka efnaskipti bæði eðlilegs brjóstaþekjuvefs og brjóstakrabbameins.

Lykilorð:

Bandvefsumbreyting, Brjóstakrabbamein, Efnaskipti, Kerfislíffræði, Mengjagögn

Abstract

Breast cancer is a heterogenous disease in which the mechanism for signalling, gene expression, metabolism and growth in breast tissue is deranged. Breast cancer is responsible for the death of over 600,000 people worldwide on a yearly basis. By investigating cellular processes in normal breast tissue, one may gain further insight into the pathophysiology of breast cancer. In this thesis, the main focal point was a developmental process called epithelial-to-mesenchymal transition (EMT), which has been shown to be fundamental for cancer dissemination. More specifically, the metabolic alterations following this process in breast epithelial cells were investigated. This thesis is in three parts. In the first part, ^{13}C isotope tracer analysis was used to characterize the metabolic differences that occur following EMT of the breast epithelial cell line D492. The results showed that following EMT, the breast cell line reduced its glycolytic activity and enhanced the reductive carboxylation of glutamine. Furthermore, the redox metabolism was affected, with glutathione synthesis being decreased significantly, which was found to affect the drug sensitivity of the cells. In the second part of the thesis, the capability of genome-scale metabolic models (GSMMs) to predict the metabolic alterations following EMT of the D492 cell line was investigated. Different types of omics data were used to constrain a breast GSMM, where the results showed that the models constrained with proteomic data were the most accurate in metabolic flux predictions when compared to results from a ^{13}C isotope tracer analysis. The proteomic GSMMs were analysed in more detail to identify metabolic vulnerabilites of the cells, which were confirmed in vitro. Finally, proteomic data from tumours of breast cancer patients were used to constrain the GSMM, which helped to identify subtype-specific prognostic markers of breast cancer patient survival. In the third and final part of the thesis, the role of the enzyme GFPT2 in EMT and breast tumourigenesis was investigated. A proteomic analysis of the D492 cell line (i) before EMT, (ii) after EMT and (iii) after overexpression of the HER2 oncogene revealed that protein levels of GFPT2 were highly correlated with the EMT status of the cell lines. Further molecular biology experiments and bioinformatic analyses revealed that GFPT2 was important for glycan metabolism and the oxidative stress response of the cells during EMT and tumourigenesis.

In summary, the results presented in this thesis increase our knowledge about the metabolic reprogramming that occurs following EMT in breast cells, and how it relates in part to the altered drug sensitivity of the cells. It was also demonstrated how useful omics data-constrained GSMMs may be in studying the metabolism of both normal and tumourigenic breast tissue.

Keywords:

Epithelial-to-mesenchymal transition, Breast cancer, Metabolism, Systems biology, Omics

Acknowledgements

I would like to begin with thanking the Icelandic Centre for Research (RANNÍS) and Göngum Saman for funding my work.

Next, I would like to thank my supervisor, Óttar Rolfsson. I first got to know Óttar when I was his teaching assistant in practical biochemistry courses at the University of Iceland from 2016-2017. In the spring of 2017, he told me that he and his colleague, Steinn Guðmundsson, were looking for a Ph.D student to model the metabolism of breast cells. The project would be an interdisciplinary one, requiring approaches used in biochemistry, molecular biology, bioinformatics, computational biology and systems biology. To my surprise, he encouraged me to apply for this position. With a background only in biochemistry, I was hesitant at first, but Óttar calmly convinced me that I had just the right experience for this project. After consulting with Margrét, who was at that time pregnant with our daughter, I decided to take a leap of faith and apply for the position. Thankfully, I got it and started at the Center for Systems Biology in May of 2017. To make a long story short, the last four years have been the best time of my life. Not only have I learnt things that I never thought I would and gotten the opportunity to develop new skills that have now been applied in published scientific studies, but I have also created an invaluable network of friends and colleagues. This time has allowed me to reflect on my passions, ambitions and interests to develop both personally and professionally. Thank you, Óttar, for being a great supervisor and a good friend.

I would next like to thank my co-supervisor, Steinn Guðmundsson, for his expert guidance and assistance during my studies. Whenever there was any problem with programming or mathemathics, Steinn would always have the answers. Steinn and Óttar make a great supervising duo, with Óttar's passion being carefully held in place by Steinn's realism.

I would also like to thank the rest of the members of my doctoral committee, Skarphéðinn Halldórsson, Margrét Þorsteinsdóttir and Þórunn Rafnar. Skarphéðinn has been a particularly important mentor during my studies. We have often engaged in academic discussions that lasted hours but felt like minutes. He provided expert scientific guidance and was a crucial part of the papers discussed in this thesis. Margrét and Þórunn have

provided expert advices and suggestions during my studies, both general and specialised related to their fields of research.

During my doctoral studies, I have gotten the opportunity to work with so many talented people. I would like to use the opportunity and mention some of them here. Arnar Sigurðsson, thanks for all the laughs and the ton of data you generated for the lab that I could use in my studies. Qiong Wang, thanks for all the discussions and collaboration over the years. Sarah McGarrity, I'm very thankful for your guidance and our discussions about metabolic modelling. I am sure that my endless amount of questions in my first year were annoying, but your answers were always helpful. Freyr Jóhannsson, your expertise in programming and metabolomics along with your advices and experiences during your graduate studies were extremely valuable for the advancement of my project. Aristoteles, your skills in everything mass spectrometry-related were helpful and our discussions about data processing and normalisation were illuminating. Adrian Garcia de Lomana, your experience and guidance proved me well in the later part of my studies. From the bottom of my heart, thank you guys. To the rest of the CSB lab, both present and past members, I am grateful for all the enlightening discussions that we had, and I am honored to have been a part of the team.

Last but not least, I would like to thank my family, to whom I dedicate this thesis. To say that graduate school is hard is an understatement. It is mentally challenging, burdening, frustrating and constantly has you questioning your own competence (impostor syndrome, anyone?). Therefore, it is crucial to have good people to rely on during the toughest times. My dearest Margrét Helga, you were always there to support me. You have had faith in my capabilities from day one and I could absolutely not have done this without you. Finally, to my kids, Ásthildur Lea and Haraldur Óli, who have been my greatest teachers, I want to say thank you for being in my life. It truly is a privilege to be your dad. I love you guys.

Contents

List of abbreviations

13C-MFA	^{13}C Metabolic flux analysis
ACN	Acetonitrile
ADP	Adenosine diphosphate
ANOVA	Analysis of variance
ASL	Argininosuccinate lyase
ASS1	Argininosuccinate synthase 1
ATP	Adenosine triphosphate
BSO	Buthionine sulphoximine
CM	Correction matrix
DDA	Data-dependent acquisition
DNA	Deoxyribonucleic acid
EMT	Epithelial-to-mesenchymal transition
ER	Estrogen receptor
ES	Enrichment score
ESI	Electrospray ionisation
FBA	Flux balance analysis
GC	Gas chromatography
GNPS	Global Natural Product Social Molecular Networking
GPR	Gene-protein-reaction
GSMM	Genome-scale metabolic model
GUK1	Guanylate kinase 1
HER2	Human epidermal growth factor receptor 2
HILIC	Hydrophilic interaction liquid chromatography
IS	Internal standard
LC	Liquid chromatography

LV	Latent variable
MDV	Mass distribution vector
MET	Mesenchymal-to-epithelial transtion
MRA	Metabolic route activity
MS	Mass spectrometry
MSoI	Metabolite set of interest
mTOR	Mammalian target of rapamycin
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PLS-DA	Partial least squares discriminant analysis
PR	Progesterone receptor
QC	Quanlity control
QC-RLSC	Quality control – Robust LOESS signal correction
RC	Reductive carboxylation
RNA	Ribonucleic acid
RPC	Reversed-phase chromatography
RPM	Reads per million reads
siRNA	Small interfering RNA
TC	Total contribution
TCA	Tricarboxylic acid cycle
TDLU	Terminal duct lobular unit
TF	Transcription factor
TGF-β	Transforming growth factor beta
TOF	Time-of-flight
UHPLC	Ultra-high-perfomance liquid chromatography
WGCNA	Weighted gene co-expression network analysis

List of figures

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-VII):

I. **Karvelsson ST**, Sigurdsson A, Seip K, Grinde MT, Wang Q, Johannsson F, Mælandsmo GM, Moestue SA, Rolfsson O, Halldorsson S., EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition. Mol Cancer Res. 2021;19(9):1546-1558. doi:10.1158/1541-7786.MCR-20-0962

II. **Karvelsson ST**, Wang Q, Hilmarsdottir B, Sigurdsson A, Moestue SA, Mælandsmo GM, Halldorsson S, Gudmundsson S, Rolfsson O. Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer. npj Syst Biol Appl. 2021;7(1):36. doi:https://doi.org/10.1038/s41540-021-00195-5

III. 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 Dec 17;21(2):100185. doi: 10.1016/j.mcpro.2021.100185.

All papers are reprinted by kind permission of the publishers.

Other publications/manuscripts:

IV. **Karvelsson ST**, Henriksen HH, Johannsson F, Madsen MB, Johansson PI, Wade CE, Rolfsson O. Taurine and lysine as metabolic markers for trauma severity. Manuscript in preparation. 2021

V. Kotronoulas A, de Lomana ALG, **Karvelsson ST**, Heijink M, Stone Ii R, Giera M, Rolfsson O. Lipid mediator profiles of burn wound healing: Acellular cod fish skin grafts promote the formation of EPA and DHA derived lipid mediators following seven days of treatment. Prostaglandins Leukot Essent Fatty Acids. 2021 Dec;175:102358. doi: 10.1016/j.plefa.2021.102358.

VI. McGarrity S, **Karvelsson ST**, Sigurjónsson ÓE, Rolfsson O. Comparative Metabolic Network Flux Analysis to Identify Differences

in Cellular Metabolism. Methods Mol Biol. 2020;2088:223-269. doi: 10.1007/978-1-0716-0159-4_11. PMID: 31893377.

VII. Wang Q, **Karvelsson ST**, Johannsson F, Vilhjalmsson AI, Hagen L, de Miranda Fonseca D, Sharma A, Slupphaug G, Rolfsson O. UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines. Mol Oncol. 2021 Dec 23. doi: 10.1002/1878-0261.13172.

Declaration of contribution

Paper I:

Skarphéðinn Halldórsson and Óttar Rolfsson designed the study. Sigurður Trausti Karvelsson, Óttar Rolfsson and Skarphéðinn Halldórsson wrote the original manuscript draft. Sigurður Trausti Karvelsson, Arnar Sigurðsson, Kotryna Seip, Maria Tunset Grinde, Qiong Wang and Freyr Jóhannsson collected the data, analysed the data and wrote the manuscript. All authors contributed to data analysis and the writing of the manuscript.

Paper II:

Sigurður Trausti Karvelsson, Steinn Guðmundsson and Óttar Rolfsson designed the study and wrote the original manuscript draft. Sigurður Trausti Karvelsson, Qiong Wang, Bylgja Hilmarsdóttir, Arnar Sigurðsson and Skarphéðinn Halldórsson collected the data, analysed the data and performed the experiments. All authors contributed to data analysis and the writing of the manuscript.

Paper III:

Qiong Wang and Óttar Rolfsson devised and planned the study and drafted the manuscript; Qiong Wang performed the experiments, analysed the data, and generated the figure plots in this study; Sigurður Trausti Karvelsson facilitated the bioinformatics analysis, including the patient survival analysis and the glutathione and GFPT2 correlation analysis; Aristotelis Kotronoulas ran the metabolomic samples with LC-MS; Þórarinn Guðjónsson provided the cell lines and helped with the functional assays; Skarphéðinn Halldórsson supervised the study and advised on both the metabolomic analysis and functional assays; Óttar Rolfsson funded and supervised this study; All authors gave their valued opinions on the manuscript.

1 Introduction

This chapter begins with a brief summary of breast cancer relevance. Then, the biology of breast gland development and epithelial-to-mesenchymal transition (EMT) will be described and how it relates to carcinogenesis. Well-known molecular drivers of EMT will then be outlined with a specific emphasis on the role of metabolism, followed by an elaborate section about metabolomics. The introductory chapter will end with describing how systems biology methods may help with analysing metabolic data and inferring metabolism from genomic data

1.1 Breast cancer

Breast cancer is a diverse disease with a prevalence of 90 cases per 100.000 people per year in Iceland (The Icelandic Cancer Society, 2021). According to the World Health Organization, it is the most common cancer worldwide and the fifth highest cause of cancer-related deaths, with 685,000 deaths in 2020 (The World Health Organization, 2021). The molecular heterogeneity of cancer is driven by various factors, including the activation of hormone receptors, growth factor receptors and loss-of-function mutations in proteins involved in the DNA damage response. To cope with the heterogeneity of breast cancer, its molecular subtyping has proven useful in the prognostic evaluation and targeted therapy of breast cancer patients. The four main breast cancer subtypes are characterised by their different gene expression profiles: Luminal A, luminal B, HER2-enriched and basal-like breast cancer (Parker et al., 2009; Perou et al., 2000). The different cancer subtypes have different levels of aggressiveness, where luminal A cancer is the least aggressive, followed by luminal B, HER2-enriched and finally basal-like breast cancer. Unsurprisingly, the aggressiveness is reflected in the survival of patients. A recent study reported that the 5-year overall survival rate of luminal A breast cancer patients was around 95% but was only 86% and 79% for patients with HER2-enriched and basal-like breast cancer, respectively (Hennigs et al., 2016). Treatment strategies for the different breast cancer subtypes depend on the tumour's histological properties and particularly on the expression of three marker receptors: human epidermal growth factor receptor 2 (HER2), oestrogen receptor (ER) and progesterone receptor (PR) (Harbeck et al., 2019). In general, the ER+/PR+ tumours (Luminal A and B) respond to hormone therapy, the ER-/PR-/HER2+ cancers (HER2-enriched)

respond to antibodies blocking the HER2-receptor (e.g., trastuzumab) but the basal-like breast cancers (ER-/PR-/HER2-) need other types of chemotherapy and radiation, where a targeted approach similar to what is used for the other breast cancer subtypes is yet to be found.

As the name indicates, breast cancer originates in the breast tissue but may eventually invade the surrounding tissue and metastasise to other parts in the body. To understand breast oncogenesis, it is important to first understand the biology of the breast gland, its origin and development.

1.2 The human breast gland

The breast (or mammary) gland is an exocrine organ responsible for producing milk in mammalian females to feed their offspring after birth. The gland has a branching structure, of which the innermost parts are called terminal duct lobular units (TDLUs) where the actual production of the milk occurs (**Figure 1**). This is also the place from where most breast cancers originate (Briem, Ingthorsson, et al., 2019). The TDLUs are formed from clusters of acini that are comprised of three cell types: (i) luminal cells, (ii) myoepithelial cells and (iii) suprabasal progenitor cells. Luminal cells are polarised and form the inner lining of the acini. Myoepithelial cells are contractile cells that surround the luminal cells and are in contact with the basement membrane. The suprabasal cells are progenitor cells with stem-cell properties that can differentiate into both luminal and myoepithelial cells (Gudjonsson et al., 2005).

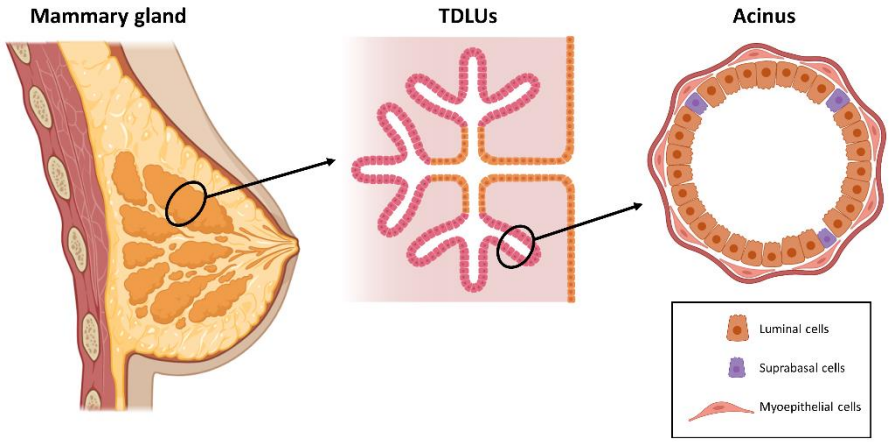


Figure 1. Schematic of the branching human mammary gland, terminal duct lobular units and acini. Adapted from Briem, Ingthorsson et al. (Briem, Ingthorsson, et al., 2019). The image icons were obtained from BioRender.

1.3 Breast development, epithelial-to-mesenchymal transition and its relevance to breast cancer progression

During embryonic breast gland development, epithelial cells migrate into the underlying mesenchyme, proliferate, differentiate and form the branch-like structures as shown in **Figure 1**. For this to happen, the polarised, adhesive epithelial cells need to alter their behaviour and become less adhesive and more mobile and invasive, which they do through a developmental process called epithelial-to-mesenchymal transition (EMT). In addition to being a part of the early embryonic development, EMT is also important during the breast gland development later in women during puberty and pregnancy, when the branch-like ductal system is expanded and becomes functional (Briem, Ingthorsson, et al., 2019). Other cell types also undergo EMT during normal biological processes, e.g., keratinocytes during wound healing. However, it is also important for various pathogenic processes like fibrosis and cancer progression (Kalluri & Weinberg, 2009).

1.3.1 Models to study the cellular & molecular biology of breast development and cancer

To study the cellular biology of the branching human mammary gland, it is crucial to have a high-quality in vitro cell culture model of the breast cellular biology. Simulating the in vivo environment of the breast cells during their culture in vitro by modifying the cell culture media and growing the cells in 3D in a synthetic extracellular matrix has enabled scientists to gain considerable insight into the cellular and molecular biology of both normal and tumourigenic breast cells (Holliday & Speirs, 2011; Hopkinson et al., 2017; Ingthorsson, Andersen, et al., 2016; Sigurdsson et al., 2011)

The D492 breast cell culture model has been used in the past decade as a tool for the investigation of the mammary gland development (Briem, Budkova, et al., 2019; Briem, Ingthorsson, et al., 2019; Gudjonsson et al., 2002; Halldorsson et al., 2017; Hilmarsdóttir et al., 2015; Ingthorsson, Briem, et al., 2016; Sigurdsson et al., 2011; Villadsen et al., 2007). It is an epithelial cell line derived from breast suprabasal cells that were obtained from a biopsy from reduction mammoplasty. The cell line was immortalised using an E6/E7 construct from human papillomavirus 16. It has been shown to have stem cell properties (Gudjonsson et al., 2002; Villadsen et al., 2007) that allow it to differentiate into both the luminal and myoepithelial cells within the acini of the breast. When co-cultured with endothelial cells in three-dimensional reconstituted basement membrane matrix, D492 cells generate branching structures that are similar to TDLUs in **Figure 1** (Briem, Ingthorsson, et al., 2019). During this co-culture, D492 cells also formed spindle-like mesenchymal colonies, from which cells were isolated to give rise to a new cell line called D492M. The D492M cells have a mesenchymal-like phenotype, indicated by their spindle-like morphology, loss of epithelial marker expression, and gain of mesenchymal marker expression (Sigurdsson et al., 2011). They also acquired a cancer stem cell-like phenotype, indicated by an increased CD44/CD24 ratio, resistance to apoptosis, increased migration and anchorage-independent growth (Briem, Ingthorsson, et al., 2019). Thus, we understand that the D492M cell line is therefore a D492 cell line that has undergone EMT. Indeed, these two cell lines have together been used as an EMT breast cell culture model in multiple studies investigating various aspects of the process (Briem, Budkova, et al., 2019; Halldorsson et al., 2017; Hilmarsdóttir et al., 2015; Sigurdsson et al., 2011). It is important to note that EMT may be induced by the overexpression of certain transcription factors, but the D492/D492M EMT model represents a natural, non-induced EMT process (although in an immortalised cell line). Other equivalent cellular

models that have been used to study EMT in breast cells include the HMLE-NAMEC model (Tam et al., 2013) and the PMC42-ET/LA model (Ackland et al., 2003).

1.3.2 Molecular drivers of EMT

Due to its ubiquity in biology, various aspects of EMT have been thoroughly investigated. The typical phenotypic changes of a cell that undergoes EMT are several. The morphological characteristics are altered by cytoskeletal rearrangements, moving from a cuboidal or squamous-like shape to an elongated, spindle-like shape. Additionally, the cell changes its behaviour, as characterised by the loss of cell-to-cell adhesion, loss of polarity, increased motility, and the ability to degrade and rearrange the extracellular matrix to enable invasion (Shibue & Weinberg, 2017). In cancers of epithelial origin (i.e., carcinomas), cells with an EMT-like phenotype have mostly been reported at the invasive front of primary tumours (Puisieux et al., 2014). Additionally, circulating tumour cells and metastatic lesions that have disseminated from the primary tumour have been shown to have EMT-like characteristics in multiple types of cancers (Pasquier et al., 2015), meaning that the EMT-derived cellular phenotype displays desirable traits for cancer dissemination and progression. In recent years it has been increasingly recognized that EMT is not a binary process, but rather a spectrum of various cellular states (Jia et al., 2021; Jolly et al., 2015; Lu et al., 2013; Pastushenko et al., 2018). The cells in different intermediate EMT stages display different proportions of epithelial and mesenchymal markers. Cancer cells with this hybrid EMT phenotype have been shown to have heightened drug resistance, stemness properties and metastatic potential (Jolly et al., 2016). This spectrum of EMT indicates that along the transition of strictly epithelial cells to become strictly mesenchymal, there are a variety of stable phenotypes that are clinically relevant.

To drive the EMT cellular phenotype, an extensive alteration in the signalling pathway activity, transcriptional pattern, proteomic composition and metabolism of the cell is required (Kalluri & Weinberg, 2009; Shaul et al., 2014). The microenvironment of cells provides stimuli that may induce EMT, that might be due to inflammation, physical stress, metabolic stress and the abnormal activation of certain signalling pathways (Puisieux et al., 2014). The signalling pathways that are known to be associated with EMT are most notably the TGF-B-SMAD, Wnt and hedgehog signalling pathways (Lamouille et al., 2014; J. Zhang et al., 2016). Alterations in cellular signalling lead to a change in transcriptional profiles of cells through the activation and inhibition

of transcription factors (TFs). EMT is carefully orchestrated by a network of TFs (most prominently the TWIST, SNAIL and ZEB TF families) which cause the change in expression of proteins and microRNAs related to cell morphology, polarity and migration, in addition to downregulation of key epithelial genes such as E-cadherin (Puisieux et al., 2014). Importantly, due to its role in cancer dissemination, the heightened expression of EMT-inducing transcription factors in tumours of various types of cancers, including bladder, breast, ovarian, colorectal, cervical and hepatocellular cancers, correlates with poor patient prognosis (Pasquier et al., 2015).

Protein level changes across EMT have been studied in various tissues. A consensus revolves around alterations in the expression of a handful of surface markers and adhesion proteins (e.g., E-cadherin, cytokeratins, vimentin, N-cadherin, fibronectin and α-smooth muscle actin), some of which are routinely screened through immunohistochemistry or Western blotting for EMT validation. However, the systematic EMT-linked changes in the proteomic composition measured by proteomic profiling and subsequent bioinformatic analysis have been gaining momentum in recent years, with proposals of novel EMT-regulating TFs, protein-protein interaction networks and functionalities (De Souza Palma et al., 2016; Grassi et al., 2017; Zhao et al., 2016).

It is well known that the metabolic phenotype of cells is altered following EMT. Yet, causal relationships between EMT and metabolism are elusive because EMT-inducing factors have been shown to both affect and be affected by metabolic changes (Jia et al., 2021). Accumulation of specific metabolites has additionally been shown to induce EMT, often by acting as cofactors/inhibitors for epigenetic regulatory enzymes, although the mechanism is not always known (Grassian et al., 2012; Sciacovelli et al., 2016; Shaul et al., 2014). Thus, the relationship between metabolism, EMT, cancer, and metastasis is still not completely understood and requires further research.

1.4 Metabolism

Metabolism refers to the set of chemical reactions that occur in organisms to produce energy, to generate building blocks for macromolecular structures and to eliminate waste products. It needs to be carefully maintained to sustain life and avoid pathological states. In the last decades, rewired metabolism has increasingly been recognised as a contributor, not a side effect, to the pathophysiology of multiple diseases including cancer (Hanahan

& Weinberg, 2011). The understanding of the metabolic rewiring that occurs during the development of pathological conditions is therefore imperative for the development of intervention strategies.

1.4.1 Metabolic alterations following EMT

The metabolic alterations following EMT have been investigated in detail in cancer cells, where aerobic glycolysis, oxidative phosphorylation, amino acid metabolism and lipid metabolism have all been shown to be affected. However, the type of change and its overall pattern relies on the tissue type and the stage within the EMT spectrum (Bhattacharya & Scimè, 2019; Jia et al., 2021). It is well known that EMT-inducing TFs SNAIL, SLUG, TWIST1 and ZEB1 all upregulate glycolysis and downregulate oxidative phosphorylation by affecting the expression of particular genes within these pathways. Yet, by directly overexpressing genes or inhibiting enzymes involved with glycolysis, fatty acid oxidation and/or glutaminolysis, it is also possible to induce EMT and metastasis in cancer cells (Jia et al., 2021).

Previously, our laboratory has created a computational model of metabolism of breast cells based on spent media measurements and transcriptomic data from D492 and D492M cells (Halldorsson et al., 2017). The analysis of these models suggested that EMT in normal epithelial breast cell line D492 reduces both glycolytic activity and oxidative phosphorylation while enhancing amino acid anaplerosis of the TCA cycle. The resulting predictions were experimentally validated by measuring oxygen consumption rate and extracellular acidification rate in both D492 and D492M cell lines. Additionally, the lipidomic profile of both D492 and D492M cells has been measured and compared in another study, where there was an overall shortening and desaturation of the fatty acid carbon chains following EMT (Eiriksson et al., 2018). Despite these findings, the metabolic phenotypes in D492 and D492M cell lines require further investigation. Further research questions would be their differential metabolic vulnerabilities and/or relationship to drug sensitivity as these could be translated directly to the dissemination of cancer cells – a clinically relevant topic.

1.5 Metabolomics

Metabolomics refers to the scientific study of metabolites within cells, tissues, biofluids or organisms. Metabolites are the end-product of regulatory processes within cells and make up the cells' metabolome (Oliver Fiehn, 2002). The metabolome directly reflects the underlying biochemical status of

cells and tissues, making metabolomics one of the leading omics technique in determining a functional phenotype. There are various groups of metabolites that are found within cells, both of endogenous and exogenous origins. These include nucleotides, amino acids, lipids, organic acids, sugars, vitamins, antibiotics and xenobiotics. These subgroups of metabolites can be thoroughly assessed separately by different sample preparation and analytical techniques (O. Fiehn, 2001), where the most widely known approaches are proteomics for proteins, glycomics for sugars and lipidomics for lipids.

There are numerous ways to acquire metabolomic data. First and foremost, there are the hyphenated mass spectrometry (MS) methods, which provide separation of metabolites within biological mixtures before being analysed in a mass spectrometer. MS is a very sensitive method for the detection and quantification of thousands of metabolites in a single analytical run. The most widely used separation methods alongside the MS are liquid chromatography (LC) and gas chromatography (GC) (Nagana Gowda & Djukovic, 2014). In addition to the hyphenated MS methods, there is nuclear magnetic resonance (NMR) which is less sensitive and less selective than the MS but for each analytical sample it is faster, requires less sample preparation efforts and offers higher reproducibility (Emwas et al., 2019)

In this chapter, the acquisition and analysis of metabolomic data will be described, which includes the type of analysis (targeted, untargeted and stable isotope-labelled), the data processing steps (normalisation and scaling methods) and the annotation of both individual metabolites and pathways. Sample extraction procedures will not be described in detail, although methods to normalise to sample handling will be discussed.

1.5.1 Mass spectrometry

Mass spectrometry measures the mass-to-charge ratio (m/z) values of ions that are derived from molecules in analytical samples. Different types of MS analysers offer unique ways of converting the ions to m/z values for quantification, where two of the most common ones are time-of-flight (TOF) and quadrupole mass spectrometry analysers (Roessner & Beckles, 2009). The TOF mass analyser accelerates the ions using an electric field and then measures the time it takes for them to reach the detector after travelling a fixed distance (**Figure 2**). Ions with lower m/z values take shorter time to reach the detector than ions with larger m/z values. Quadrupole mass analysers are composed of four cylindrical rods that are in parallel with a

fixed space in between them. A fixed voltage is applied between the rods with a direct current (DC) offset, forming an oscillating electric field in the space between the rods. The ions travel through the quadrupole alongside the rods, but only ions within a narrow range of m/z values will reach the detector. Other ions will have unstable trajectories and collide with the rods before reaching the detector. The quadrupole can therefore be thought of as both a mass analyser and an ion filter (Roessner & Beckles, 2009).

Many laboratories use two or more mass analysers in tandem for mass spectrometry (also called MS-MS or MS^2). It is common to use a quadrupole first as a filter to select only ions within the mass-to-charge range of interest followed by molecular fragmentation. The fragmentation can be achieved through several methods, e.g., collision-induced dissociation, electron capture dissociation or photodissociation. The fragmentation pattern depends on the molecular structure of the ions and is highly reproducible. Finally, the ions (fragmented or not) go into another mass analyser, which can be a quadrupole (TQMS) or a TOF (qTOF-MS). This approach of using a series of mass analysers is termed tandem-mass spectrometry and is used to identify compounds (or *analytes*) in analytical samples with high confidence by comparing the fragmentation patterns to those of known standards (Roessner & Beckles, 2009).

1.5.2 Separation techniques for metabolites

In order to obtain further resolution of the metabolome within complex biological mixtures, mass spectrometry is often coupled with a separation technique that occurs prior to the MS analysis. This can be achieved through chromatography, which separates metabolites based on their physicochemical properties. Chromatography consists of two phases, a mobile phase and a stationary phase. The mobile phase is the phase which dissolves the biological mixture and carries it through the stationary phase which interacts differently with metabolites in the mixture based on their properties. Metabolites with high affinity for the stationary phase take a longer time to move through it than metabolites with lower affinity which forms the basis for the mixture's separation.

We can distinguish chromatographic methods based on the physical state of the mobile and stationary phases. When the mobile phase is liquid and the stationary phase is solid, the technique is called liquid chromatography. In the field of metabolomics, liquid chromatography coupled with a mass spectrometer (LC-MS) has become one of the most widely used methods

due to its broad coverage of polar and non-polar metabolites and lack of requiring any chemical derivatization steps (Nagana Gowda & Djukovic, 2014). Chemical derivatization of metabolites is often required for other types of chromatographic methods, like gas chromatography, which has a gaseous mobile phase and a solid stationary phase. In liquid chromatography, the analytical sample is injected into the stream of the mobile phase that is then transported under pressure onto the column containing the solid phase. When the pressure of the mobile phase is high (100-300 bars), the technique is called high-performance liquid chromatography (HPLC), and when the pressure is very high (over 1,000 bars), the technique is called ultra-high-performance liquid chromatography (UHPLC).

Common types of liquid chromatography used in metabolomics include the hydrophobic interaction liquid chromatography (HILIC) and the reversed-phase chromatography (RPC). HILIC is when the stationary phase is hydrophilic, meaning that it is coated with polar molecules (e.g., amides and amines) that interact with polar analytes dissolved in the mobile phase during the chromatographic run. This type of chromatography is suitable for the separation of polar analytes (e.g., amino acids, organic acids and sugars). At first, the HILIC mobile phase is non-polar (e.g., 100% acetonitrile/ACN) to ensure that the hydrophobic analytes are dissolved and pass through the LC column first. The polarity of the mobile phase is then increased by adding a polar solvent (e.g., water) in increasing amounts, slowly increasing the solubility of polar analytes and disturbing their interaction to the stationary phase, eventually leading to their elution from the column. In contrast, RPC has non-polar alkyl-chains in its stationary phase, meaning that it is ideal for the separation of non-polar analytes. The mobile phase system works in opposition to the one for HILIC, with the polarity of the mobile phase being high at first and lowered following the chromatographic run to allow for the elution of hydrophobic analytes.

1.5.3 Interface between liquid chromatography and mass spectrometry

In LC-MS, the analytes eluting from the column are dissolved in the liquid mobile phase, so they need to be (i) converted into the gaseous phase and (ii) ionised. Both can be achieved simultaneously using a soft-ionisation technique termed electrospray ionisation (ESI) (Fenn et al., 1990). In the ESI interface, or the *ion source*, the LC-eluting analytes are nebulised with the help of a nebulizing gas while going through a high-voltage metal capillary where charged droplets are formed. The charged droplets enter an ionisation

chamber where they are evaporated with the help of a heated desolvation gas and/or heating of the capillary, leaving only the ionised analytes that subsequently enter the mass analyser with the help of focusing voltages (Pitt, 2009). The metal capillary is usually orthogonal to the inlet of the mass analyser to avoid contamination. The voltage polarity of the metal capillary and ion optics can be tuned so that only positive or negative ions are measured which is important as the ionisation of the analytes depends on their functional groups. A simplified schematic of an LC-MS with an ESI interface is shown in **Figure 2**.

In positive ionisation mode, the ions that are typically formed during ESI are protonated analytes $[M+H]^+$, whereas in negative mode they are deprotonated analytes $[M-H]^-$. When conducting measurements in positive mode, the protonation of analytes can be significantly enhanced by lowering the pH of the mobile phase in the LC method (e.g., by adding formic acid). In contrast, when conducting measurements in negative mode, the deprotonation of analytes may be enhanced by increasing the pH of the mobile phase (e.g., by adding sodium bicarbonate). These adjustments to the mobile phase, that cause some analytes to be already ionised in the liquid phase, may increase the ESI-MS sensitivity significantly. However, the resolution of the LC chromatographic separation may be decreased as the analytes' ionisation is enhanced, suggesting that compromises must be made (Kostiainen & Kauppila, 2009). In addition to the protonated and deprotonated analytes, other ion derivatives, or *adducts,* can form, such as $[M+Na]^+$, $[M+ACN+H]^+$, $[M+NH_4]^+$ in positive mode and $[M-H_2O-H]^-$, $[M+Na-2H]^-$ and $[M+Cl]^-$ in negative mode. The adduct formation generally depends on the solvent of the analyte, constituents of the mobile phase and the glassware (Kruve & Kaupmees, 2017).

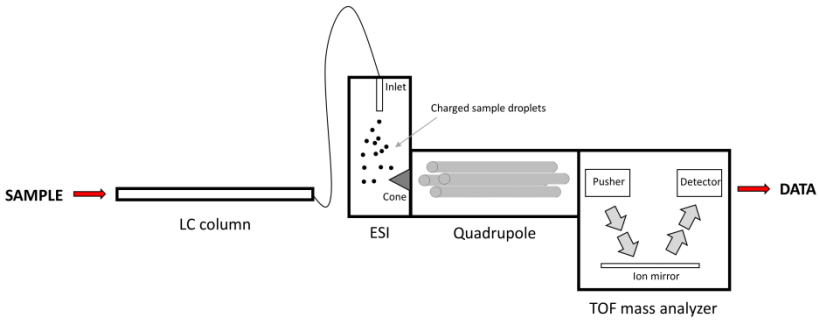


Figure 2. Schematic of qTOF LC-MS with an ESI interface. The sample is injected into the liquid chromatography (LC) column, where metabolites elute at different timepoints based on their physicochemical properties. The metabolites are ionised at the electrospray ionisation (ESI) interface, after which they enter the quadrupole, which serves as a mass filter to select ions within a certain m/z range. The selected ions enter the TOF mass analyser where the ion signal intensity is measured to produce data with three dimensions: 1) retention time, 2) m/z value and 3) signal intensity. The inlet in the ESI interface represents the high-voltage metal capillary where the sample is converted to gaseous phase and ionised. The cone in the ESI interface represents the inlet to the quadrupole which is orthogonal to the ESI metal capillary inlet.

1.5.4 Types of metabolomic data analysis

The data generated in a typical metabolomic analysis using the hyphenated MS methods contains thousands of measurements which represent the plethora of metabolites within the metabolome of the analytical sample. Furthermore, molecule fragmentation in the ion source and adduct formation cause many metabolites to be represented by multiple datapoints. The data for a single LC-MS run is usually three dimensional, containing information about the mass-to-charge ratio, retention time and the signal intensity of detected ions (**Figure 3**).

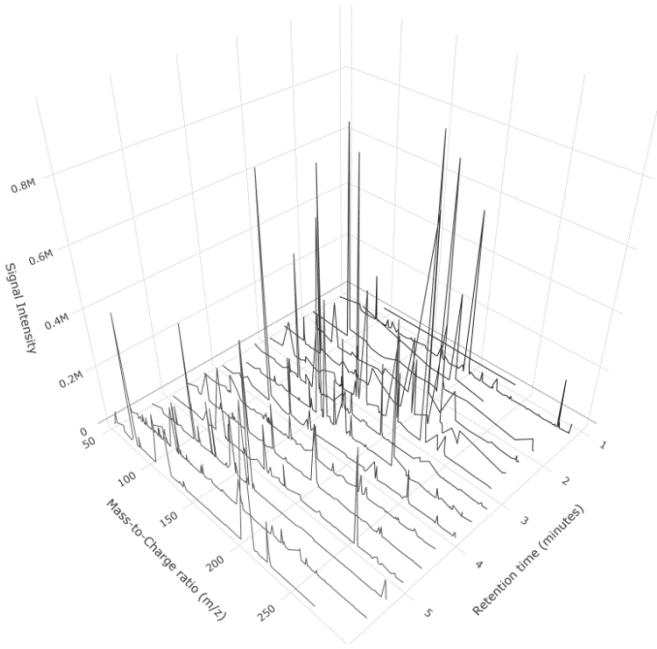


Figure 3. The three dimensions of LC-MS data. This 3D density plot visualises the results from a single analytical sample from an LC-MS run. Each peak represents a single feature (m/z) value eluting at the given retention time window (in this case 20 seconds). The signal intensity represents the ion abundance which is measured in the MS detector.

At this stage, the detected and quantified ions are referred to as features, as they have not yet been annotated as particular metabolites. When running multiple analytical samples there is need for specialised software for data processing, i.e., for noise filtering, feature detection, alignment of features between samples, normalisation of features, removal of batch effects and downstream analysis (Katajamaa & Orešič, 2007). There are two main approaches used for the analysis of metabolomic data: Targeted and untargeted.

1.5.4.1 Targeted metabolomic analysis

A targeted analysis of LC-MS data is when detected features are compared to an in-house library containing the m/z values and retention times of known metabolites that have been thoroughly characterised in the same system using MS^2 and individual and/or mixtures of metabolic standards (**Figure 4**).

The usage of standards allows scientists to account for variations brought by the *matrix effect* and sample handling, and to quantify the absolute concentrations of metabolites of interest.

The matrix effect is when the ionisation of an analyte of interest is affected by the co-elution of other compounds, both endogenous and exogenous, from the LC. The matrix effect depends on the LC-MS method, the analyte of interest and the ion source and may lead to either ion enhancement or ion suppression, both leading to the incorrect quantification of the analyte of interest (N. Fang et al., 2015). To account for the matrix effect, the analytical samples are often spiked with a mixture of *internal standards* (ISs) before extraction, where the concentrations of the various metabolites are known. This mixture ideally contains a fixed concentration of stable isotope-labelled analogues of the analytes of interest or other compounds that have near-identical physicochemical properties. A stable isotope-labelled analogue can be distinguished from its analyte analogue in an extracted ion chromatogram. In theory, the matrix effect on the analytes of interest and their IS counterparts should also be near-identical, meaning that the matrix effect can be quantified due to the known concentrations of the various IS. In addition to accounting for the matrix effect, the IS mixture may also be used for the correction of variations introduced by sample handling and processing during the extraction procedure (S. Wang et al., 2007). After adjusting the analyte signals to internal standards, the data are now considered *semi-quantitative,* meaning that although the absolute concentrations of the analytes of interest are still unknown, their relative abundances in different analytical samples correspond to the actual differences present between samples (Liu & Locasale, 2017).

External standards (also known as reference standards) have mostly been used for the absolute quantification of analytes of interest. Within a single analytical run, standard solutions with varying concentrations of specific analytes of interest (external standards) are measured alongside the samples that need quantification (unknown samples) within the same analytical run. A calibration curve is generated for the analytes of interest within the standard solutions based on the signal intensity values and their known concentrations. This is achieved using a line of best fit, resulting in a linear equation for the calibration curve. The absolute concentrations of the analytes of interest in the unknown samples are then calculated based on the equation that describes the calibration curve.

1.5.4.2 Untargeted metabolomic analysis

In contrast to a targeted analysis, an untargeted metabolomic analysis includes measurements of all features above a fixed peak intensity threshold, resulting in a global metabolic profile of the analytical sample, along with all the other data processing steps. It does not rely on previous measurements performed in the same system. Furthermore, it opens the possibility to generate data-driven hypotheses (**Figure 4**). This approach requires different computational tools than a targeted approach. A broadly used tool is the open-source software package XCMS (Smith et al., 2006), which can be used for automatic chromatographic peak detection and retention time correction (i.e., alignment of analytical runs). Other similar tools are MS-DIAL (Tsugawa et al., 2015), MAVEN (Clasquin et al., 2012) and MZmine (Pluskal et al., 2010), in addition to the software dedicated to untargeted analysis provided by most mass spectrometry vendors. Untargeted metabolomic analyses are usually semi-quantitative, since it is practically a daunting task to account for all features using specific external standards. An important step prior to any type of analysis of untargeted metabolomic data is the removal of variability introduced by the LC-MS platform or the personnel facilitating the sample preparation steps. This procedure will be outlined in the following section.

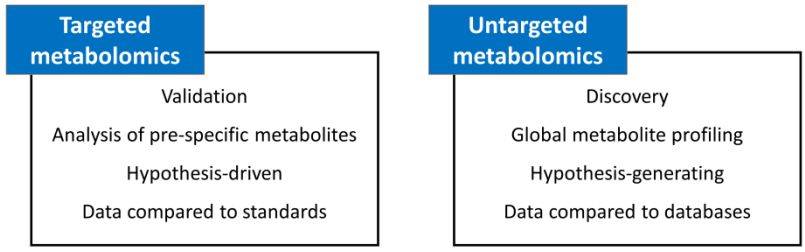


Figure 4. Targeted and untargeted metabolomic analyses. The two different approaches used in metabolomics are targeted and untargeted analyses. Both possess different traits shown in the separate boxes. Adapted from Schrimpe-Rutledge et al. (Schrimpe-Rutledge et al., 2016).

1.5.4.3 Processing of semi-quantitative metabolomic data

In all metabolomic analyses, there are multiple factors that can introduce noise or cause drift in signal intensities both within analytical batches and between them. This unwanted systematic variation needs to be removed before any further processing of the data. Thus, before starting any LC-MS

analytical runs, the experiment needs to be planned thoroughly to eliminate all possible sources of unwanted variation. A typical analytical run starts with the conditioning of the LC column with a sample containing small amounts of all analytical samples - a *pooled* sample. Then the solutions containing the external standards are run, followed by the samples that are run in a randomised order with the pooled samples being run at regular intervals. The run ends by measuring the external standards again. The pooled samples can act as quality-controls (QC) and should theoretically give the same total-ion chromatogram in each run. This information can be used for normalisation purposes, and is the foundation for the quality control – robust LOESS signal correction method (QC-RLSC) (Dunn et al., 2011) which normalises the intensity of the sample features to a non-linear locally estimated smoothing function that is fitted to the pooled QC samples' features.

Another normalisation method is the aforementioned usage of internal standards. As previously mentioned, the IS mix should ideally contain a paired IS for each analyte in the analytical samples. The drawback of using IS for normalisation is that internal standards are not available for all analytes and they can be very expensive (Stokvis et al., 2005). Therefore, while IS normalisation is a suitable technique in targeted metabolomic analyses, normalising all features using ISs in untargeted metabolomics is very challenging in practise. It is possible to normalise features to an IS that does not have the same molecular formula, although this may lead to additional errors if the representative IS is not sufficiently similar in structure and/or behaviour (Gullberg et al., 2004). However, one can use multiple internal standards to both normalise each feature to circumvent a possible cross signal contribution between the IS and the analytes (Redestig et al., 2009) and find the optimal normalisation factor for each feature (Sysi-Aho et al., 2007).

After the normalisation steps, the metabolomic dataset consists of intensity values and retention time for all features in all samples of the analytical run, similar to the three-dimensional plot in **Figure 3** but for all samples. The intensity values for the different metabolites (or features, depending on whether one has identified the metabolites) may differ by orders of magnitude and their intensity distribution in the analytical samples is often skewed (van den Berg et al., 2006). Before moving on to downstream analysis, the metabolomic data need to be scaled and standardised so the unit of variance is identical between features. This is so the features can be statistically compared between samples and to facilitate gradient descent and

distance-based algorithms. If the data have a heavily right-skewed distribution of feature intensity values, they require some sort of a power transformation (e.g., log, square root or generalised log (Durbin et al., 2002)) to approximate a normal distribution before the scaling and standardisation steps. The method chosen to scale the data can significantly affect the results of the analysis (van den Berg et al., 2006). Among the most common scaling methods are autoscaling, pareto scaling and variance-stabilizing (vast) scaling, each with their own advantages and disadvantages.

1.5.5 Annotation of metabolomic data

A crucial step in untargeted metabolomics is the identification, or *annotation*, of metabolites within the dataset of measured features. As the features only contain information about the mass-to-charge ratio, the retention time and the signal intensity (**Figure 3**), the chemical identification of LC-MS data can be particularly challenging. The general way of identifying metabolites in analytical runs is by running multiple stable isotope-labelled chemical standards and/or using tandem mass spectrometry (MS^2) to compare the fragmentation patterns of unknown features to other patterns in a spectral library, e.g., online or in-house. Data-dependent acquisition (DDA) is an approach used in untargeted metabolomics, where the MS^1 spectra are measured to provide accurate quantification of the overall intensity of all detected features, followed by the fragmentation and measurement (MS^2) of the features with the highest signal intensity to aid with their identification. In this way, the MS^2 spectra for the features with the highest intensity are collected automatically when the MS^1 intensity exceeds a predefined threshold (Schrimpe-Rutledge et al., 2016). The DDA is usually conducted for a subset of the samples, or the pooled samples, and can be used later during the metabolite annotation step by comparing the fragmentation spectra of the features of interest to those in spectral libraries. Many laboratories have their own in-house spectral libraries for different LC-MS methods which require considerable amount of time and resources to establish. Therefore, the identification of metabolites is frequently a bottleneck in the functional interpretation of metabolomic analyses (Dunn et al., 2013).

When only considering the mass-to-charge ratios, each m/z value can map to multiple different metabolites, which may or may not have the same structural formula. Additionally, adduct formation and molecular fragmentation during the ionisation step in LC-MS expands the list of possible m/z matches even further. Adding the retention time information enhances the resolution of the LC-MS data, since compounds with similar structures

and functional groups tend to coelute from the LC column. However, even including the retention time information may not resolve compounds with the same m/z values that behave similarly, e.g., stereoisomers, which may require further efforts, e.g., specialised LC-MS methods or NMR (Dunn et al., 2013). When MS^2 spectra from analytical samples are available (e.g., after conducting a DDA), the annotation of metabolic features may become easier. The Global Natural Product Social Molecular Networking (GNPS) infrastructure has in recent years been expanding as a tool for the molecular networking of MS^2 data and their annotation (Aron et al., 2020). The MS^2 fragmentation spectra of the unknown samples are compared to those within a public data repository (GNPS-MassIVE) containing hundreds of thousands of reference MS^2 spectra.

Once metabolomic data has been acquired, normalised, scaled and annotated (if desired), the next steps depend on the experimental hypothesis. For a comparison of individual metabolite abundance between two or more analytical samples, univariate analysis is used. To visualise metabolic fingerprints in a more collective manner, multivariate analysis is used.

1.5.6 Univariate metabolomic analysis

For comparison of individual metabolite levels between groups, parametric tests, such as Student's t-test and analysis of variance (ANOVA), are used when the metabolomic data are normally distributed (possibly after nonlinearly transforming the data), and the comparison groups have similar data variance. If the data fails to meet these criteria, non-parametric tests are used (e.g., Mann-Whitney U-test for two groups and Kruskal-Wallis test for more than two groups).

When conducting multiple statistical tests, as usually is the case for metabolomic data containing hundreds to thousands of features, one must adjust for multiple comparisons to reduce the number of false positives (Type I error). There are various methods available for p-value adjustment, such as the Bonferroni correction which controls the family-wise error rate, where the chosen significance threshold (e.g., $\alpha = 0.05$) is divided by the number of tests (m) to generate a new significance threshold (α/m). Another commonly used method is the Benjamini-Hochberg method (Benjamini & Hochberg, 1995), which controls the false-discovery rate (FDR). This method is less conservative than the Bonferroni method (Korthauer et al., 2019).

1.5.7 Multivariate metabolomic analysis

Metabolomic data typically contains measurements of multiple metabolic features (k) which usually outnumber the analytical samples (n). Thus, if all features of metabolomic data were to be used in some type of model describing a particular phenotypic trait, it would typically result in non-unique model parameter estimates which may be overfitting the data (Johnstone & Titterington, 2009). A convenient way of avoiding overfitting data in models is using regularisation techniques. Regularisation causes the penalisation of non-zero model coefficients that ultimately reduces variance at the cost of increasing bias in predictions. However, in addition to the "large k, small n" problem, metabolomic data is often characterised by multicollinearity, which is when numerous metabolic features have highly correlated signal intensities. This collinearity might be due to adduct formation or fragmentation in the ion source during an LC-MS analysis, or simply that two or more metabolites are very similar in behaviour. Therefore, even if the features used in a model would be fewer than the analytical samples ($n > k$), collinearity needs to be addressed beforehand or the estimates of the model coefficients may be imprecise.

In a targeted metabolomic analysis, collinearity is addressed by simply removing one of two highly correlated features. In an untargeted metabolomic analysis, where the features are yet unidentified, two or more features that display collinearity may be useful for the annotation and functional interpretation of the data (see section 1.5.8), so removing any of them may interfere with that process. Other feature eliminating methods that retain the original properties of the data but still reduce its dimensionality are required. The choice of dimensionality reduction technique depends on the experimental hypothesis. For a completely exploratory analysis, principal component analysis (PCA) is often chosen as the first step. Briefly, PCA finds the multiple perpendicular axes along the metabolomic datapoints which preserve most of the variation in the data. These axes, called eigenvectors, are as many as the features within the metabolomic data. By ranking the eigenvectors with respect to their eigenvalues which correspond to explained variance, most of the variance is often preserved within the first few eigenvectors. The original metabolomic data is then transformed to a new subspace by projecting it onto the first few eigenvectors explaining most of the variance, forming the *principal components* (PCs) (Worley & Powers, 2013). The data can now be visualised along the principal components of interest to observe if there is any group structure present and whether it relates to any known class labels of the analytical samples. This method has

both been used to drive discovery or for quality assurance of hyphenated MS analyses (Eiriksson et al., 2018; Kotronoulas et al., 2020a; X. Li et al., 2019; Rusilowicz et al., 2016).

Partial least squares discriminant analysis (PLS-DA) is another commonly used dimensionality reduction technique. Unlike PCA, PLS-DA is a supervised method where the known class labels of analytical samples are used to determine an optimal linear transformation that separates the different classes of samples (Worley & Powers, 2013). It projects the original metabolomic data to latent variables (LVs) so that the discrimination between the two classes is maximised. An extension of this method is orthogonal projection to latent structures (OPLS-DA) which performs the additional step of removing systematic variation that is unrelated to the analytical sample classes (Worley & Powers, 2016).

A quantified phenotypic trait can be regressed on selected PCs or LVs directly. Furthermore, the metabolic features that correlate with the PCs or LVs of interest can be identified and used directly for further analysis – a process called feature selection. When it comes to choosing between PCA and (O)PLS-DA, knowing the class labels and finding the latent variables that define the separation between two or more classes in a metabolomic dataset may seem like an obvious choice. However, by successively adding Gaussian noise to a metabolomic dataset, Worley and Powers found that OPLS-DA will still identify differences between classes even when there effectively is none (Worley & Powers, 2016). The authors suggested that a PCA should be used for qualitative assessment of the reliability of supervised models applied to metabolomic data, meaning that the PCA and PLS-DA approaches should be used together.

1.5.8 Functional interpretation of metabolomic data

While quantitative analysis of the metabolome can provide insight into the metabolic fingerprint of a specific cell, tissue or biofluid, it cannot directly report on the activity of individual reactions or pathways. For a direct measurement of the individual activity of enzymes there are specific assays available for various reactions. However, repeating these assays for tens or hundreds of reactions in various pathways is time-consuming and may not provide optimal results as the experimental conditions may vary. In order to systematically estimate metabolic activity from metabolomic data, there are a variety of methods available. The most widely used method is metabolite set analysis, which requires results from a single “metabolic snapshot” where two

or more conditions are compared. Other methods rely on either measurement of the uptake and secretion rates of multiple metabolites or the incorporation of stable isotopes into metabolites of interest (see section 1.5.9).

1.5.8.1 Metabolite set analysis of metabolomic data

Metabolite set analysis takes place after obtaining results from univariate or multivariate metabolomic analysis. The analysis generally assesses whether there is a significant enrichment of, or impact on, predefined, functionally related metabolites. Functionally related metabolites refer to metabolites that may be in the same metabolic pathway, belong to the same tissue location or are known to be associated with a specific disease. These predefined metabolites will now be referred to as metabolite set of interest (MSoI). There are three varieties of metabolite set analysis that are most commonly used: metabolite set enrichment analysis, overrepresentation analysis and pathway impact analysis.

Metabolite set enrichment analysis uses the information from a complete metabolomic dataset (i.e., whose dimensionality has not been reduced) to identify whether there is a significant enrichment of particular MSoI (Xia & Wishart, 2010). It is essentially a metabolic version of the popular gene set enrichment analysis (Subramanian et al., 2005). Concisely, the input is a ranked list of all metabolites where both the quantity and direction of the statistical difference is captured. Then, for a MSoI, an enrichment score (ES) is calculated from the ranked list and compared to a null distribution of ES obtained from a permutation of the list of metabolites. An empirical p-value is calculated by comparing the observed ES to the null distribution of ES from the permutation procedure. The p-values are then typically adjusted for multiple hypothesis testing (Subramanian et al., 2005).

Overrepresentation analysis is a method that does not require any ranking of the inputs. For a subset of metabolites of size *k* from a complete metabolic dataset of *K* features (e.g., all metabolites with a p-value below the specific significance threshold from a univariate analysis), the observed number of metabolites related to a particular MSoI is obtained. Then the expected number of metabolites from the same set of interest is calculated when selecting *k* random metabolites from the complete metabolite dataset of *K* features. The observed and expected ratio of metabolites within the MSoI are statistically compared using a Fisher's exact test, hypergeometric test or a chi-squared test (only for larger set sizes).

Pathway impact analysis is an extension of the general overrepresentation analysis where the topological properties of prespecified MSoI are incorporated (Xia et al., 2011) to calculate an impact that a list of input metabolites has on those MSoI. Briefly, an importance measure, which can be the centrality or degree, of all metabolites in a network constructed from the complete metabolomic dataset is calculated within each MSoI. Then, for the list of input metabolites, a pathway impact is calculated as the sum of the matched metabolite importance measure, divided by the total importance measures within the MSoI. Topological network properties are described in more detail in section 1.8.

The above algorithms are only useful when the metabolomic data is annotated. However, in the past decade, considerable efforts have been made to circumvent the metabolite identification bottleneck by performing automatic annotations of detected features in hyphenated MS^1 analyses (Alonso et al., 2011; Kuhl et al., 2012; Senan et al., 2019; Uppal et al., 2017). A few algorithms, notably the mummichog algorithm (S. Li et al., 2013), skip the identification of individual metabolites and use instead the collective information of multiple m/z values to associate the metabolite alterations to metabolic pathways (**Figure 5**). In short, the mummichog algorithm takes a list of m/z values (e.g*.,* from all features below a certain p-value threshold, or the features with the highest loadings from a PCA) and sequentially identifies all possible matches for each m/z value (within a narrow range) from an online database. A permutation test is then used to identify whether there is a significant localised enrichment of mapped metabolites within specific metabolic pathways. Importantly, this procedure takes only a few minutes at most, depending on the parameter settings. In comparison, the manual annotation of a complete untargeted metabolomic analysis may take weeks to months to complete. The mummichog algorithm is especially useful for a rapid functional interpretation of high-throughput metabolomic datasets (Hagan et al., 2019; Sinclair et al., 2017), informing hypotheses that ultimately require confirmation using a targeted metabolomic analysis.

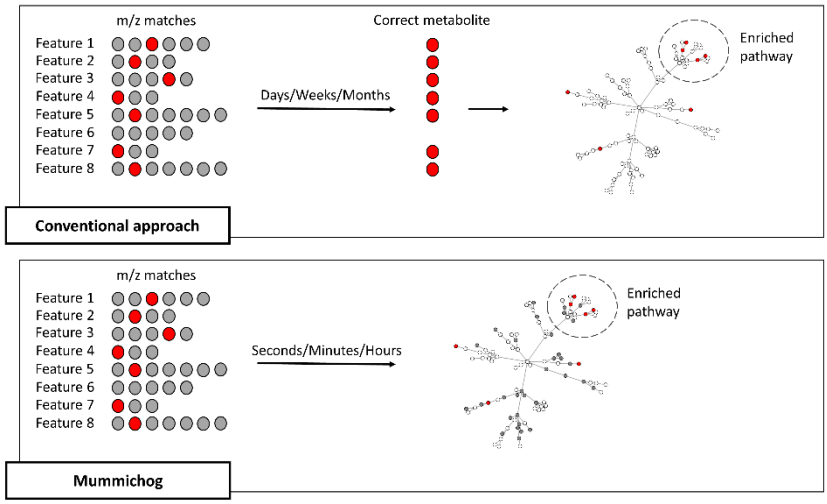


Figure 5. Graphical comparison of conventional metabolomiic annotation and the mummichog algorithm.The two approaches have a list of eight features that they want to map to a metabolic network. In the conventional approach, there is a lot of time spent on the manual annotation of features based on m/z values, retention time, reference standards, MS^2 spectra etc. This results in a small group of correctly annotated metabolites that are mapped to a metabolic network, where there is significantly enriched mapping of the feature list to a specific pathway. In the mummichog approach, all possible m/z matches in an online database (e.g., the Human Metabolome Database (Wishart et al., 2018)) are identified for the complete list of features. All the putative matches are mapped to the metabolic network, where there is a significantly enriched mapping of the feature list to a specific pathway. Adapted from Li et al. (S. Li et al., 2013).

1.5.9 Stable isotope-labeled metabolomic analysis

Targeted and untargeted metabolomics can provide snapshots of the metabolome. A series of metabolic snapshots over time can be used to measure the net uptake and secretion rates of metabolites in cells in culture which can then be used to infer metabolic activity. However, changes in intracellular metabolite concentrations over time does not allow us to infer anything about metabolic rates, or *fluxes*, of individual intracellular reactions because these are a part of a larger metabolic network. For example, in **Figure 6**, an increase in the concentration of metabolite B does not tell you whether it is due to increased activity of reaction 1, decreased activity of reaction 2 or some combination of both.

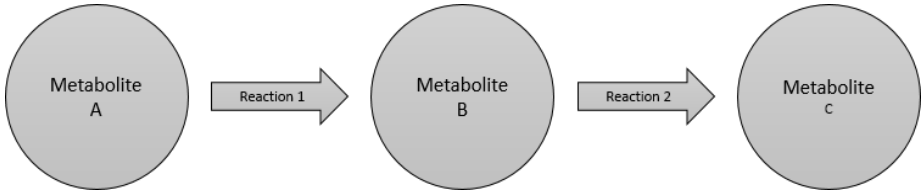


Figure 6. A diagram showing a metabolic pathway comprised of three metabolites (circles; A, B and C) and two reactions (arrows; 1 and 2).

One way of inferring metabolic reaction and pathway activity in cells is by feeding them carbon sources that are labelled with stable isotopes, e.g., ^{13}C, ^{15}N and ^{2}H. After a certain amount of time in culture, the isotope will be incorporated into all metabolites that derived from the selected carbon source, causing multiple isotopologues of each metabolite to be formed within the cell. Isotopologues are molecules that differ only in their isotope composition (see **Figure 7** for details). After quantification of each isotopologue, the isotope labeling patterns, or the mass distribution vectors (MDV), of the metabolites can be calculated. An MDV represents the relative abundance of the different isotopologues of a certain metabolite and can be used to infer metabolic activity (Buescher et al., 2015). For example, in an experiment where cells have been fed with a fully ^{13}C-labelled carbon source, a metabolite that contains *n* carbons can have 0 to *n* carbons that are ^{13}C-labelled. The MDV of this metabolite is calculated based on the relative ion count of each isotopologue compared to their total combined ion count as measured by a mass spectrometer (**Figure 7**).

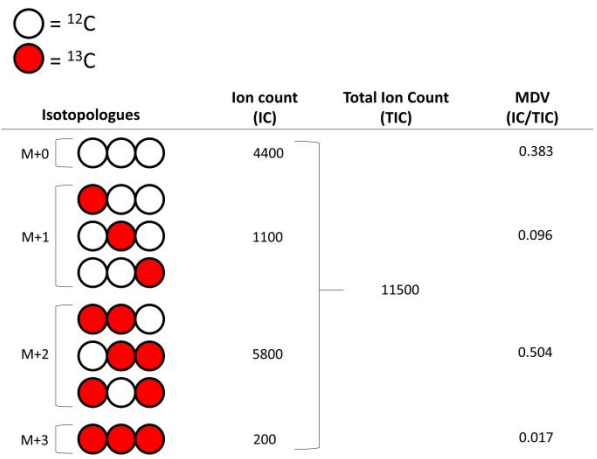


Figure 7. Example of calculation of the MDV vector from the ion count/intensity values of the different isotopologues of a 3-carbon metabolite. The ion count are a hypothetical MS-derived quantification of the different isotopes that are used to calculate the hypotheital MDV. The exact position of the ^{13}C isotope within the carbon skeleton of isotopologues cannot be deciphered using MS, but rather with NMR.

Importantly, before the calculated MDV can be used for further analysis, it needs to be corrected for naturally occuring isotopes. For many atoms, there are two or more stable isotopes that occur naturally. For example, carbon has both ^{12}C and ^{13}C which have 98.93% and 1.07% natural abundance, respectively. Other carbon isotopes are radioactive and undergo exponential decay which can be exploited in age determination of objects.

Thus, to quantify observed isotope labels in a metabolite from a stable isotope carbon source, the relative abundance of naturally occuring isotopes needs to be taken into account. This is done using correction matrices (CM) for each element in the metabolite of interest with naturally occuring isotopes. The corrected MDV is calculated as follows:

$$MDV_{cor} = CM^{-1} \cdot MDV_{meas} \quad (1)$$

Where CM^{-1} is the inverse of the correction matrix based on the abundance of naturally occuring isotopes of all elements within the metabolite of interest, MDV_{cor} is the corrected MDV vector and the MDV_{meas} is the measured MDV vector as shown in **Figure 7**. If the metabolite contains only carbons, oxygen and hydrogen atoms, then the CM is calculated as follows:

$$CM = CM_C \cdot CM_O \cdot CM_H \quad (2)$$

An example of a CM for a molecule with *n* carbons is shown in the equation below:

$$CM_C = \begin{pmatrix} {}^{n}C_{12} & 0 & \cdots & 0 \\ (n-1)C_{12} \times 1C_{13} & {}^{n}C_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ {}^{n}C_{13} & 1C_{12} \times (n-1)C_{13} & \cdots & {}^{n}C_{12} \end{pmatrix} \quad (3)$$

Where C_{12} and C_{13} represent the natural abundance of the ^{12}C and ^{13}C isotopes, respectively. The non-zero values, or *combined abundances*, in equation 3 are calculated using the following equation (Van Winden et al., 2002):

$$Combined\ abundance = \left(\sum_{i=1}^{N} f(I(i))\right)! \cdot \prod_{i=1}^{N} \left(\frac{p\big(I(i)\big)^{f(I(i))}}{f\big(I(i)\big)!}\right) \quad (4)$$

Where the element of interest has *N* naturally occurring isotopes, *I(1)* to *I(N),* with natural abundances *p(I(1))* to *p(I(N))* and frequencies *f(I(1))* to *f(I(N))* within the isotopologue. The correction for naturally occurring isotopes is usually performed using a specific software such as IsoCor (Millard et al., 2019).

After correcting for the naturally occurring isotopes, what remains are the MDV_{cor} vectors for all the metabolites of interest. These can be compared between cells, treatments or timepoints to get a relative difference in the labelling profiles. A convenient way of comparing and presenting the MDV vectors is by using the total contribution (TC), also called mean enrichment. It represents the total contribution of a stable isotope-labelled carbon source to a metabolite of interest, and only requires the MDV_{cor}. It is calculated as follows:

$$Total\ contribution = \frac{\sum_{i=1}^{n} i \cdot m_i}{n} \quad (5)$$

Where, for all *n* isotopologues, a weighted average of their corrected, relative abundance (*m*) is calculated. Higher TC values mean that there is higher net flow of stable isotopes into the metabolite of interest, meaning that there must be higher activity in the pathway leading to the formation of said

metabolite. Statistical tests can be used to compare the TC values between groups.

1.6 Systems biology

In the biological sciences, there are two main paradigms – reductionism and holism. The former refers to the detailed investigation of individual components, whereas the latter refers to the analysis of a whole system of multiple components, their interactions and emergent properties, i.e., properties of a system as a whole that cannot be observed in the individual nodes (Loscalzo & Barabasi, 2011). Although the reductionist approach has been and will continue to be indispensable for generating novel information about various biological phenomena, there has been a paradigm shift in recent years within the research community (Palsson, 2015). The emergence of high-throughput omics methods has enabled scientists to generate immense amounts of data in a short amount of time. These large datasets often contain the simultaneous measurements of multiple variables ($n = 10^2$ to 10^4), making it difficult to stick to the reductionist approach. This escalation in data size and complexity has led scientists to alter their philosophy to a systems-level thinking of how the biological components interact globally. This approach in biological sciences has been termed *systems biology*.

In this section, two common systems biological approaches will be presented. Methods based on the topological properties of networks in the analysis of metabolomic data will be described, and then the usage of genome-scale metabolic models for metabolic activity inference will be discussed.

1.6.1 Topology analysis of biological networks

A system of components and their interactions is often formalised as a network. In a network, individual components are nodes and their interactions are edges. For example, in a network representing metabolism, the nodes would be metabolites and the edges would be metabolic reactions. Another example is the human protein interactome (Luck et al., 2020), where proteins are nodes and edges between nodes exist if those proteins have been shown to interact. Network-based approaches have proven useful in analysing networks of multiple individual components, since diseases are rarely the consequence of a single flawed gene or a protein, but rather the dysregulation of a large, intracellular and/or intercellular network within multiple tissues and organs (Barabási et al., 2011). The biological networks

can be characterised and analysed using techniques and applications from graph theory (Koutrouli et al., 2020). One important network property is its degree distribution. The degree of a node represents the number of its connections to other nodes, i.e*.,* its number of edges. In various biological networks the degree distribution has been shown to follow a power law, where the network is said to have a scale-free (SF) topology (Jeong et al., 2000; Koutrouli et al., 2020) (**Figure 8A**). However, the SF topology of biological networks is highly controversial (Broido & Clauset, 2019; Tanaka, 2005), and it has been suggested that other distributions (e.g., exponential or log-normal) could fit the degree distributions of biological networks equally well as, or even better than, the power law distribution (Broido & Clauset, 2019; Lomana et al., 2010). Regardless of the best fitting distribution, biological networks seem to have in common that their degree distributions are heavily right-tailed, meaning that most of the network nodes have very few interactions, but there are a few nodes that have a very high connectivity (**Figure 8B**), i.e*.,* many edges or interactions, and are sometimes referred to as hubs. In biological networks, there is typically a high degree of modularity, which means that there are multiple subsets of nodes within the network that are highly interconnected but sparsely connected to other nodes (or other node subsets) as observed in **Figure 8C**. Modules typically contain one or more hubs that have the highest connectivity values among the nodes in the module. Hub removal usually has the broadest impact on the network functions (Palsson, 2015).

Another important node property is their betweenness centrality (**Figure 8D**). It is based on a network property called the shortest path, which is the path between two nodes in the network that contains the least amount of edges. If the edges in the network are weighted, then the shortest path is the path where the sum of the edge weights between two nodes is minimised. For a particular node *k*, the betweenness centrality is an enumeration of the shortest paths between all pairs of nodes in the network that pass through *k,* normalised to the total number of shortest paths in the network (Koutrouli et al., 2020). Nodes that have the highest betweenness centrality values form so-called bridges, or bottlenecks (Barabási et al., 2011), between communities and have been shown to be important for the connectivity of the network (Koschützki & Schreiber, 2008).

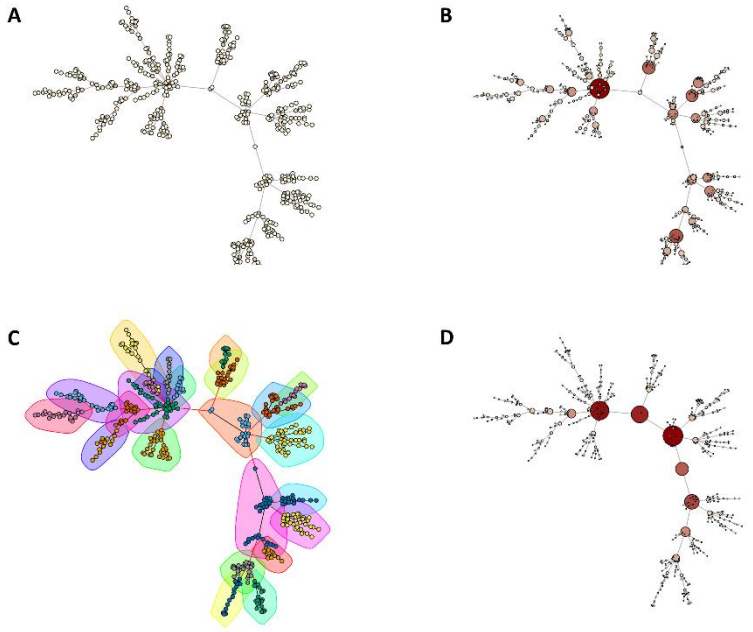


Figure 8. Visualisation of network properties. A scale-free network was constructed using the Barabasi-Albert model (Albert & Barabási, 2002) with n = 500 nodes. The Fruchterman-Reingold force-directed layout (Fruchterman & Reingold, 1991) was used to visualise the network. A) The network. B) The network where both node sizes and intensity of red colour represent their degree. C) The network with highlighted modules identified using the Louvain method for community detection (Blondel et al., 2008). D) The network where both node size and intensity of red color represent their betweenness centrality.

The connection between nodes does not have to be physical. Gene co-expression networks are constructed based on the gene-gene correlation within datasets. In this case, the edges are said to be weighted, where the weight is the absolute correlation value. The modules within these networks are often genes that are regulated by the same transcription factors or belong to the same biological processes/pathways (Farhadian et al., 2021; Niu et al., 2019; Yang et al., 2018; Z. Zhang et al., 2020). The modules can only be identified after constructing a gene co-expression network and are useful for investigating the functionality of genes, which would not have been possible using typical reductionist approaches.

In addition to revealing the emergent properties of a system, network analysis also allows the dimensionality reduction of biological data, since the modules may be summarised into a single vector through singular-value

decomposition (Langfelder & Horvath, 2007). Network topology analysis is therefore a versatile systems biological method that allows the functional interpretation, target identification and dimensionality reduction of biological data.

1.6.1.1 Network medicine

The application of network-based approaches to the study of human disease is defined as *network medicine* (Barabási et al., 2011; Chan & Loscalzo, 2012; Silverman et al., 2020). Network medicine uses interaction networks between molecular and environmental components in human cells to reveal drivers of disease. The overall goal is to use innovative network-based techniques to analyse big data, and translate the discoveries to the clinic to improve disease diagnosis, prevention and treatment (Parini et al., 2020). This often involves the identification of disease modules, i.e., neighbourhoods of components that contain relatively many disease-related components. A typical example is the usage of the human protein interactome along with a list of genes or proteins known to be affected by a certain disease (Gysi et al., 2021; Menche et al., 2015). This list of genes/proteins may be acquired from online databases such as OMIM and UniProt (Hamosh et al., 2005; Menche et al., 2015; Mottaz et al., 2008), by statistical tests or some other means. The list of genes/proteins is mapped to the human protein interactome, and the topological relationship of the list components is evaluated. Chances are that the genes/proteins in the list are highly interconnected, or a specific module of the interactome is enriched with members of the gene/protein list. Either way, this leads to the identification of a neighbourhood (or neighbourhoods) within the interactome which is now called the disease module (or modules). The disease modules may then be tested for overlap with other disease modules, which may aid with the disease characterisation or drug development, if the overlapping disease is well characterised and is known to be affected by certain drugs. In this way, network-based approaches may be used for repositioning of already approved drugs instead of relying on conventional drug development procedures, thus minimizing cost, time and risk (Parini et al., 2020).

A bibliometric analysis of the PubMed database using the Dimensions database (Herzog et al., 2020) in September 2021 revealed that there are over two thousand studies describing the usage of network topology for drug repurposing, and the number of additional studies is increasing every year. Despite some compelling proposals of drug repositioning, many of the studies are limited by the lack of experimental data to confirm the effects of

the proposed drugs. For example, in the midst of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) global pandemic, there were hundreds of network analysis papers published that described drug repurposing for treatment of patients affected by the coronavirus. Most of them proposed targets that needed further verification, but a few had some level of verification by screening infected cell lines or retrospectively investigating COVID-19 clinical trials. Gysi and colleagues found that out of the 200 top-ranked potential drug repurposing candidates against SARS-CoV-2, only 13 had positive outcomes in primate cells infected with the virus, of which 6 had strong effects on infected human cells (Gysi et al., 2021). These drugs were auranofin, fluvastatin, azelastine, digoxin, methotrexate and vinblastine, of which the first three have been associated to mitigating effects on SARS-CoV-2 infections in either independent cell screenings or retrospective clinical studies (Konrat et al., 2020; Marić et al., 2021; Rothan et al., 2020). In a similar COVID-19 study, there were also multiple false positives among the main predicted drugs (Zhou et al., 2020), which raises the question of whether these types of network analyses are worthwhile compared to well-organised, high-throughput drug screenings for the identification of effective drugs.

1.6.1.2 Network topology analysis of metabolomic data

Although the scale-free features of metabolic networks have been questioned, a heavy-tailed degree distribution and presence of hubs (e.g*.,* ATP, ADP, acetyl-CoA) is a definite property of metabolic networks (Takemoto, 2014). Metabolites participating in the same metabolic pathways can be expected to have higher correlation in abundance than metabolites in different metabolic pathways (Rosato et al., 2018), meaning that the topology of a metabolic network may be reconstructed based on a metabolite-metabolite correlation matrix from a metabolomic analysis (Arkin et al., 1997; Fukushima et al., 2011; Jahagirdar et al., 2019; L. Wang et al., 2015). This may prove particularly helpful when analysing an untargeted metabolomic dataset where the features' identity remains unknown, where the modular organisation of the correlation network may aid with annotation and functional interpretation of the data, e.g*.,* via the *mummichog* algorithm (S. Li et al., 2013), where the input list of features is a whole module (**Figure 5**). In addition, dimensionality reduction may be achieved through this approach as previously described. However, despite these useful insights brought by the topological properties of metabolic networks, they alone are not enough to assess the functional state of the network, that is to say the metabolic activity.

For organism-scale metabolic activity inference and assessment of how the metabolic activity is related to the organism's genotype, the usage of genome-scale metabolic models (GSMM) has proven practical. GSMMs will now be described and their application in health sciences highlighted.

1.6.2 Genome-scale metabolic models

The genome of an organism is the complete set of its genetic information, providing all the necessary information for the organism to survive in the appropriate environment. The genome contains all the organism's genes, including both protein-coding and non-protein-coding genes. Genes code for RNA transcripts, the complete set of which it is called the transcriptome. After RNA splicing and maturation, transcripts may be translated into proteins, the complete set of which is called the proteome.

Genome sequencing is the determination of the sequence of adenine (A), thymine (T), cytosine (C) and guanine (G) in the genome at a single point in time. The sequencing of the human genome was finished in 2003 as a result of the combined efforts of multiple laboratories worldwide. This collaboration, called the Human Genome Project, took 13 years to complete which highlights the size and complexity of the *Homo sapiens* genetic blueprint. Despite that the official sequencing effort was finished almost two decades ago, the functional annotation of the human genome is still an ongoing project (Frankish et al., 2019). This annotation refers to the interpretation of the DNA sequence, i.e., finding gene locations and coding regions and determining their roles.

How does a genome define metabolic functionality, and to what degree? A genome defines which genes and proteins, including enzymes, are available for the cell to use, meaning that it provides the framework for the metabolic functionality of the cell. The genome and its annotation can be used to construct genome-scale metabolic models for cells and organisms, where all metabolic genes, their enzyme product and the associated reactions/metabolites are defined. Even more biochemical data can be incorporated, such as metabolite and protein structure (Brunk et al., 2018). However, as described in section 1.3, EMT is a complex biological process where a cell undergoes phenotypic changes, altering their morphology, behaviour, signalling, gene expression, protein composition and metabolism. Importantly, the cell's genotype is not affected by this process, implying that the same genotype may account for multiple metabolic phenotypes. Therefore, the GSMM of an organism needs to be tailored to a specific

phenotype through omics integration, which creates context-specific GSMMs that describe the metabolic functionality on a phenotype-specific level.

1.6.2.1 Reconstruction of GSMMs

For a bottom-up reconstruction of a GSMM, the first thing required is an annotated genome. Thiele and Palsson break the construction of GSMMs into four stages (Thiele & Palsson, 2010).

The first stage is automatic creation of a draft reconstruction. This is achieved by using gene ontology categories to specifically extract metabolic genes from the annotated genome. This will also have information about the protein products (enzymes) and functions of genes, which can otherwise be obtained from a biochemical database like the Kyoto Encyclopedia of Genes and Genomes (KEGG). The second stage is manual refinement of the genes within the metabolic reconstruction, representing the comprehensive re-evaluation of all extracted metabolic genes, their enzyme products and function. It includes the assembly of the reactions into pathways, the verification of each reaction's substrate and cofactor usage, stoichiometry and directionality, re-evaluation of the gene-protein-reaction (GPR) rules which indicate which gene encodes what protein and elicits what function, determining metabolite charges at given pH based on pKa values, adding necessary transport reactions, and determining the biomass composition, growth-associated ATP maintenance reaction and requirements from a growth medium. This is the most time-consuming part of all the reconstruction process. The third stage is the conversion of the draft reconstruction to a mathematical format, followed by the fourth stage, network evaluation, which mainly revolves around testing the reconstruction's ability to produce biomass along with other predefined, necessary metabolic tasks depending on the organism being modelled. If the task testing fails, the troubleshooting steps include gap analysis, removal of thermodynamically infeasible loops, testing for production of biomass function components, and comparing predicted physiological properties with known properties. If the troubleshooting steps are unsuccessful in resulting in a functional reconstruction, stages 2 – 4 are repeated as needed. The complete reconstruction process can take months to years to complete (Thiele & Palsson, 2010).

The first human GSMM, Recon 1, was reconstructed in 2007 and was based on an annotated human genome and extensive literature evaluation (Duarte et al., 2007). Later in the same year, another group published the Edinburgh human metabolic network reconstruction, also based on genomic

and bibliomic data (Ma et al., 2007). The merging of the two models along with additional manual curation, gap-filling and consistency tests resulted in Recon 2 in 2013 (Thiele et al., 2013). The most comprehensive human metabolic reconstruction to date, Recon3D, was published in 2018 and contains over thirteen thousand metabolic reactions and four thousand metabolites (Brunk et al., 2018). The human metabolic reconstructions represent the metabolic capabilities of most cells within the human body. However, cells in different tissues do not display the same metabolic phenotypes, which means the human metabolic reconstructions need to be tailored to generate a context-specific GSMM representing the cell/tissue of interest. This is the basis behind the various GSMM model building algorithms that are available (e.g,, INIT (Agren et al., 2012), mCADRE (Y. Wang et al., 2012), FASTCORE (Vlassis et al., 2014) and CORDA (Schultz & Qutub, 2016)). These model building algorithms integrate a generic GSMM and context-specific experimental/literature data, where the goal is to create a consistent, context-specific subnetwork from the generic GSMM, in a relatively short amount of time. This is done by either a top-down or a bottom-up approach where reactions are either removed or added, respectively, based on their score within the experimental data. A systematic evaluation of the methods for data integration with GSMMs revealed that no single model building algorithm is superior to the rest at generating models capable of predicting the activity of metabolic reactions in central carbon metabolism (Machado & Herrgård, 2014). However, a study of multiple model-building algorithms for extracting functional metabolic models from the generic human reconstruction Recon 1 to represent various cancer cell lines showed that the choice of model-building algorithm had the greatest impact on the accuracy of predicting essential metabolic genes. Furthermore, the authors suggested that the inclusion of a core set of reactions and known tissue-specific metabolic functionalities improves the accuracy of the models (Opdam et al., 2017).

1.6.2.2 Constraint-based analysis of GSMMs

Cells operate under constraints, both physicochemical and environmental, in addition to the self-imposed constraints produced by regulation, captured by their transcriptome and/or proteome (Palsson, 2015). The physicochemical constraints are hard, non-adjustable constraints that result from the natural laws (e.g*.,* conservation of mass, reaction stoichiometry, osmotic pressure, pH, solvent capacity and molecular diffusion). While biology obeys the natural laws, the function of a biological system cannot be predicted by them alone due to the unpredictable constraints brought on by evolution, which are time

variant. However, physicochemical constraints provide the range of the possible behaviours of biological systems, which can be further limited by the environmental conditions (e.g*.,* the concentration of particular metabolites in the extracellular environment of a cell), transcriptional regulation, and abundance of enzymes, all of which are adjustable constraints. Therefore, when modelling the functionality of metabolic systems, successive applications of different constraints result in a set of possible functional states, or a *solution space,* of metabolic systems rather than a single functional state. Each point in the solution space represents an allowable functional state that obeys the applied constraints. Importantly, when the environment, transcriptional profile or protein/enzyme abundances of a system change, the constraints are altered and subsequently the allowable functional states of the system (Palsson, 2015).

When referring to the modelling of a functional state of a metabolic system, it means obtaining the rates of the reactions of the system, or the change in its metabolite concentration values over time. The reaction rates are also termed flux values, and the vector of all reaction rates of a system a flux vector. The reaction stoichiometry is usually described in a matrix *S,* where the columns represent reactions and the rows represent metabolites. The entries in the matrix are the stoichiometric coefficients of the reactions in the system.

In a metabolic system, $\mathbf{v} = (v_1, v_2, \ldots v_n)$ is a flux vector for all *n* reactions, and $\mathbf{x} = (x_1, x_2, \ldots x_m)$ is a concentration vector of all *m* metabolites. The stoichiometric matrix *S* transforms the flux vector into a vector of time derivatives of metabolite concentrations (**Equation 6**):

$$\frac{d\boldsymbol{x}}{dt} = S\boldsymbol{v} \tag{6}$$

In constraint-based analysis of metabolic systems, it is common to assume that the system is at steady-state due to the lack of kinetic parameters for metabolic reactions at the genome-scale. This means that the metabolite concentrations remain constant over time:

$$S\boldsymbol{v} = \mathbf{0} \tag{7}$$

Metabolic systems usually have more reactions than metabolites, meaning that they are under-determined (Orth et al., 2010). Therefore, when a metabolic system is constrained only by the steady-state assumption, it has an infinite number of solutions confined in an n-dimensional hyperplane

solution space. The solution space can be reduced to a convex polytope by adding further constraints, not only physicochemical ones (e.g., reaction directionality based on Gibbs free energy change), but also environmental ones (e.g., uptake and secretion rates of extracellular metabolites), and regulatory ones (e.g., maximum and minimum bounds on reactions based on the enzyme abundance or activity). Once the solution space has been defined (**Figure 9**), it can be analysed in order to obtain biologically relevant functional states of the metabolic system.

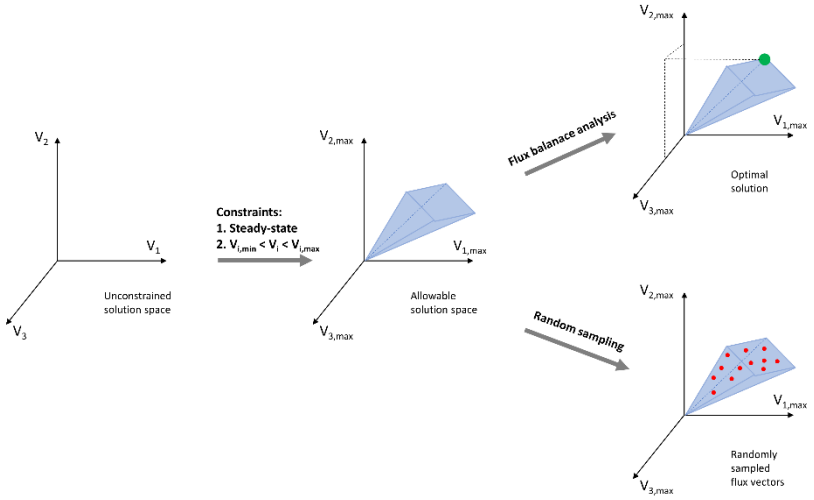


Figure 9. Constraint-based modelling and analysis. Before any constraints are added, the model has an infinite number of solutions. After the addition of the steady-state assumption and reaction upper and lower bounds, the solution space becomes a convex polytope. The solution space can be investigated using both flux balance analysis, in which an optimal solution (indicated by a large green dot) is found for the maximisation of a particular function (in this case, V_2). In random sampling, flux vectors that are within the allowable solution space are sampled (indicated by small red dots). Adapted from Fang et al. (X. Fang et al., 2020).

The most widely known method for functional analysis of constraint-based metabolic models is flux balance analysis (FBA). This method uses linear optimisation to identify the maximal (or minimal) value of a predefined objective function within the model (**Figure 9**) satisfying all the constraints. The output is a flux vector for all reactions in the model, but with the maximum (or minimum) possible value for the objective function (Orth et al., 2010). The objective function of the model may be ATP production, biomass production or any other combination of reaction fluxes that the researcher is interested in. However, identifying the objective of organisms is not always straightforward. In many prokaryotes, the objective is often assumed to be

growth as a product of evolution, and therefore biomass production is a reasonable objective. In multicellular organisms, the metabolic objective of individual cells depends on their location and developmental stage which is not always known and is not necessarily biomass production or ATP production, but may also be the optimisation of cofactor balances in different subcellular compartments (García Sánchez et al., 2012; Zielinski et al., 2017) or the simultaneous optimisation of multiple tasks (Dai et al., 2019).

Random sampling of the solution space identifies thousands of random functional states that are within the allowable solution space (**Figure 9**). It results in a distribution of flux values for each reaction that can be compared between differently constrained models or to identify reaction sets that are highly correlated (Schellenberger & Palsson, 2009). Unlike FBA, there is typically no optimisation procedure involved with random sampling where a certain function is being maximised, so it identifies (mostly) flux vectors that are sub-optimal. This can be particularly useful in identifying the systemic metabolic effects brought on by enzymopathies (Price et al., 2004) or different physiological conditions (Thiele et al., 2005).The main shortcoming of this method is that it is time-consuming, although different sampling procedures have recently been introduced to address this problem (Haraldsdóttir et al., 2017; Megchelenbrink et al., 2014).

To summarise, the extracellular metabolome (measured over time to quantify uptake/secretion rates), transcriptome and proteome of cells can be used to create context-specific models of metabolism through integration with GSMMs. These are the adjustable constraints, where under certain circumstances (e.g*.,* following a developmental process such as EMT) the abundance of certain extracellular metabolites or certain enzymes is altered which should be reflected in the solution space of the GSMM.

1.6.2.3 GSMMs and human disease

GSMMs represent the metabolic networks of cells/tissues/organs and therefore provide a convenient platform for the integration of disease-related omic data to evaluate the systemic metabolic effects of said diseases. Considerable effort has been made to model the metabolism of human diseases using GSMMs, including different pathologies of the liver (Bidkhori, Benfeitas, Elmas, et al., 2018; Bidkhori, Benfeitas, Klevstig, et al., 2018; Björnson et al., 2015; Shubham et al., 2017), lung cancer (Asgari et al., 2018) prostate cancer (Marín de Mas et al., 2018) and endothelial dysfunction (McGarrity et al., 2018). Agren and colleagues predicted 101 antimetabolites (i.e., drugs that inhibit the use of specific metabolites through competitive

inhibition) that could be used to treat patients with hepatocellular carcinoma, 22 of which are currently used in cancer therapy (Agren et al., 2014). Mardinoglu et al. showed that by constructing and analysing personalised GSMMs for patients with non-alcoholic fatty liver disease, they were able to identify differences in glutathione and NAD^+ metabolism in patients with a high degree of fatty liver, which they confirmed using proof-of-concept studies in mice and human subjects (Mardinoglu et al., 2017). More recently, by integrating cell line-specific transcriptomic data, metabolomic data, ^{13}C isotope tracing data, oxygen consumption rate measurements and data from high-throughput siRNA screening results with a human GSMM, Tarrago-Celada and colleagues found that metastatic colorectal cancer cells have exploitable vulnerabilities in folate and cysteine metabolism, which they confirmed in vitro (Tarragó-Celada et al., 2021). Overall, the construction and analysis of human disease GSMMs has in recent years resulted in clinically relevant proposals of metabolic vulnerabilities, drug repositioning and patient stratification. With increasing quality of human genome annotation and availability of different types of high-throughput omics data, the constraint-based analysis of context-specific GSMMs of human diseases will undoubtedly continue to add mechanistic insight into their pathophysiology and propose vulnerabilities that may be exploited in their treatment. What is currently lacking is a definitive consensus in the type of data and model construction procedures that provide the most accurate predictions.

2 Aims

Although the metabolism of EMT in breast cells (both tumourigenic and non-tumourigenic) has been intensively investigated during the last decade, there is still debate about how the various EMT-inducing TFs affect metabolism and how metabolism, in return, affects the transcriptional landscape of the cells (Bhowmik et al., 2015). The general aim of this project is to further elucidate the metabolic alterations that occur following EMT in normal human breast tissue using metabolomics and systems biology approaches. There was special emphasis on the potential usage of transcriptomics and/or proteomic data to infer metabolism since these are the readouts of the genome (genotype), whereas metabolomics is the functional readout of the cell (phenotype). This would help bridging the gap between genotype and phenotype.

There are three specific aims in this project:

1. To define the contribution of different carbon sources to the metabolism of breast tissue cell lines before and after EMT: This will be done using the D492 and D492M cell culture model of EMT and isotopic nutrient labelling,

2. To define the metabolic phenotypes of EMT across genome-scale models (GSMMs) by integrating existing breast cell GSMMs with different types of omics data and assess which type of omics data best capture EMT-linked metabolic alterations.

3. To predict metabolic vulnerabilities on a cancer subtype-specific level using breast GSMMs and omics data.

3 Materials and methods

In this chapter, I describe the main methods used in the studies outlined in this thesis. Additional description of materials and methods can be accessed in the corresponding chapters of the papers presented in section 4.

3.1 Cell culture

D492 and D492M cells (with and without *IDH2*-shRNA construct) were cultured at 37°C and 5% CO_2 in H14 medium. H14 is DMEM/F12 supplemented with 250 ng/ml insulin, 10 μg/ml transferrin, 10 ng/ml EGF, 2.6 ng/ml sodium selenite, 10^{-10} M oestradiol, 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 hours. For the isotope labelling experiments, cells were grown until 80% confluent in normal H14 medium, at which point the medium was changed to H14 medium containing 100% ^{13}C-labeled glutamine at the 1 or 5 position (Cambridge Isotope Laboratories, Inc., MA, USA) or ^{13}C-labeled glucose at the 1 and 2 positions (Cambridge Isotope Laboratories, Inc.). Cells were cultured in the labelled medium for 6 hours, at which point the sample extraction procedure was executed.

The D492 and D492M cell lines were kindly provided by the Stem Cell Research Unit, BioMedical Center, University of Iceland and were screened for Mycoplasma infection monthly by the BioMedical Center, University of Iceland.

3.2 Metabolomics

3.2.1 Sample extraction

Methanol/chloroform/water extraction was used to extract metabolites from cell cultures. Briefly, cells were harvested using ice-cold methanol, vigorously mixed using a vortex mixer and left on ice for 10 minutes. Then, equal amounts of water and chloroform were added to a obtain a final composition of 1:1:1, after which samples were mixed using a vortex and left to stand overnight at 4°C. The organic phase was first collected into a glass tube and dried in a stream of gaseous nitrogen and stored at -80°C until the next steps in the LC-MS analysis. The aqueous phase was stored at -80°C and evaporated in a miVac concentrator (SP Scientific, Warminster PA, USA) after which it was ready for LC-MS analysis.

3.2.2 UPLC-MS

The organic phase was reconstituted in methyl tert-butyl ether (MTBE) before a methanol solution containing 1M NaOH was added (10:1 v/v, respectively) before the samples were incubated for 210 minutes at 37°C, after which 1 µL formic acid was added, the samples were dried in a stream of gaseous nitrogen and resuspended in isopropanol:ACN:H_2O (6:9:1, v/v/v). The aqueous phase (metabolites) was reconstituted directly in isopropanol:ACN:H_2O (2:1:1, v/v/v).

For the LC-MS analysis, an ultra-performance liquid chromatography (UPLC) (Acquity, Waters, Manchester, UK) was coupled with a quadrupole-time of flight mass spectrometer (Synapt G2, Waters, Manchester, UK). For the organic phase samples, chromatographic separation was achieved as previously described (Kotronoulas et al., 2020b). For the aqueous phase samples, chromatographic separation was achieved by hydrophilic interaction liquid chromatography (HILIC) using an Acquity amide column, 1.7 µm (2.1 x 150 mm) (Waters, Manchester, UK). Samples were run under acidic and basic chromatographic conditions in both positive and negative ionisation modes. In the acidic conditions, mobile phase A was 100% ACN and B was 100% H_2O both containing 0.1% formic acid. The following elution gradient was used: 0 min 99% A; 7 min 30% A; 7.1 min 99% A; 10 min 99% A. In the basic conditions, mobile phase A contained ACN:sodium bicarbonate 10 mM (95:5) and mobile phase B contained ACN:sodium bicarbonate 10 mM (5:95). The following elution gradient was used: 0 min 99% A; 6 min 30% A; 6.5 min 99% A; 10 min 99 % A. In all conditions, the flow rate was 0.4 mL/min, the column temperature was 45°C, and the injection volume was 3.5 µL. A capillary voltage of 1.5 kV was used to operate the mass spectrometer, whereas the sampling cone and the extraction cone were of 30 V and 5 V, respectively. The cone and the desolvation gas flow were 50 L/h and 800 L/h, respectively, while the source and desolvation gas temperature were 120°C and 500°C, respectively. The MS spectra were acquired in centroid mode from m/z 50 to 1000 using scan time of 0.3 s. As lock mass, 2 ng/µL leucine enkephalin was used (m/z 556.2771 and 554.2615 in positive and negative modes respectively). A typical analytical block consisted of (i) pooled QC samples for system equilibration, (ii) calibrators containing varying concentrations of standards, (iii) analytical samples and spiked pooled QC samples and (iv) calibrators again.

3.2.3 Data analysis

For the integration of the chromatograms of the metabolites of interest, TargetLynx (v4.1; Waters) was used. Ion chromatograms were extracted using a window of 0.02 mDa which was centred on the expected m/z for each targeted isotopologue. The output was a mass distribution vector (MDV) describing the relative amount of each detected isotopologue of the metabolite. The MDVs were corrected for abundance of naturally occurring isotopes using the IsoCor software (Millard et al., 2012). To summarise the ^{13}C label incorporation into metabolites, the mean enrichment or total contribution (TC) was used (Buescher et al., 2015) which was calculated using **Equation 5**.

3.3 Constraint-based metabolic modelling

Genome-scale metabolic model construction and analysis was carried out in MATLAB using the COBRA Toolbox (Heirendt et al., 2019). A previous genome-scale breast tissue metabolic model (Halldorsson et al., 2017) was used as a base model, which was constructed by tailoring of the generic human metabolic reconstruction Recon 2 (Thiele et al., 2013). Briefly, based on an RNA sequencing (RNA-seq) dataset from both D492 and D492M, all genes with ≥ 1 RPM in either cell line were selected and their associated reactions was identified (through Recon 2's Gene-Protein-Rules). The identified reactions were used as the set of "core reactions" which is one of the required inputs for the FASTCORE model building algorithm (Vlassis et al., 2014), which was subsequently used to construct the context-specific metabolic network for D492 and D492M cells. This network was manually curated to ensure no major metabolites or pathways were missing or blocked, resulting in the breast metabolic model referred to as iBreast2886.

3.3.1 Construction of breast epithelial and mesenchymal GSMMs

The iBreast2886 model was used to create cell-type specific models of epithelial D492 and mesenchymal D492M based on microarray, proteomic and RNA-seq data. A detailed workflow description of the model construction is included in the Supplementary information in Paper II. 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). An additional 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. Thus, in the end there were 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 for the MES models.

3.3.2 Analysis of breast epithelial and mesenchymal GSMMs

Random sampling of the solution space was used to estimate the distribution of possible flux vectors of all EPI and MES models. The random sampling results were used for two different analyses.

First, the median value for the flux distribution of each reaction was used to represent its activity for all reactions in the models, resulting in a "median flux vector". This flux vector was then used as a representative flux vector for the models, for comparing one model to another (Supplementary figure 1, Paper II) and for the estimation of the minimal alterations in reaction bounds that are required for EPI models to take on the MES median flux vectors under the steady state assumption. The latter is based on an algorithm previously developed in our laboratory (Choudhary et al., 2016) to solve the following optimisation problem:

$$minimize \ \|v - v_{MES}\|_1 \quad (8)$$

Where $\boldsymbol{v}$ are the decision variables, $\boldsymbol{v_{MES}}$ is the median flux vector of MES and $\|.\|_{\mathbf{1}}$represents the Manhattan norm (1-norm). The list of reactions with altered bounds was then tested for the over-representation of reaction subsystems (or metabolic reactions families/pathways) within iBreast2886.

Secondly, we compared the results of random sampling to ^{13}C isotope tracing experiments using a metabolic route activity (MRA) measure which we specifically designed for the task to estimate the relative contribution of extracellular metabolites to intracellular metabolites of interest in all randomly sampled flux vectors. The MRA is outlined in detail in Supplementary figure 6, Paper II. Briefly. in order to calculate the relative flux value v_{rel} from metabolite m_i to m_{i+1} within a pathway of interest in a particular GSMM, we first identified all consuming reactions of metabolite m_i using the stoichiometric matrix S. Then, for each individual flux vector, we calculated the total consuming flux of metabolite m_i, and the ratio of the total flux used to produce only metabolite m_{i+1}, which we will call v_{rel}. Let k be 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 shown in Equation 9:

$$v_{rel}(k) = \frac{v(k)}{\sum_{i=1}^{K} v(i)} w_{comp} \quad (9)$$

Where w_{comp} is the compartmental weight given to the relative flux value based on volume of the compartment it occurs in, as some reactions occur in multiple cellular compartments. For all transport reactions, the $v_{rel}(k)$ values were defined as 1.0. The compartmental weights were based on their relative volume within a typical liver cell (Johnson et al., 2009), where the values were as follows: Cytosol 0.54, mitochondria 0.22, ER 0.12, nucleus 0.06, Golgi apparatus 0.03, peroxisomes and lysosomes 0.01.

It is possible to calculate the MRA for a predefined subset of metabolites that form a path, or a route, in the GSMM network. To obtain the MRA from metabolite m to $m + k$ in a particular route, we calculate the sum of the log of relative flux values (from **Equation 9**) within that route:

$$MRA = \sum_{i=1}^{k-1} \log(v_{rel}(m_{i+1})) \quad (10)$$

Where the first reaction is the consumption of metabolite m_i to produce metabolite m_{i+1}. We performed the MRA calculations for all randomly sampled flux vectors in each GSMM to get a distribution of MRA values for each GSMM.

In addition to random sampling, we used FBA in the single gene knockout simulations to estimate whether the maximum biomass production rate is affected when all associated reactions of a specific gene (through the GSMM‘s GPRs) have their bounds set to zero. This was performed for all genes within a GSMM using COBRA Toolbox‘s singleGeneDeletion function (Heirendt et al., 2019) to identify the essential genes for that particular GSMM, i.e., the genes whose deletion has an inhibitory effect on the maximum biomass production rate.

3.4 Drug-drug correlation network analysis

In the work presented in this thesis, a correlation network analysis was used for the grouping of drugs based on their effects on the cell lines within the NCI-60 Human Tumour Cell Line panel (Shoemaker, 2006). We gathered the drug sensitivity data from the NCI-60 Human Tumour Cell Lines using the

rcellminer R-package (v. 2.10.2) (Luna et al., 2016), where we focused specifically on FDA-approved drugs. After filtering the data, the final size of the drug sensitivity matrix was 214 drugs x 59 cell lines (the cell line MDA-N was removed due to missing data). We employed the WGCNA R-package (v. 1.70.3) (Langfelder & Horvath, 2008) to construct a network of drug-sensitivities, where we used a soft threshold value of 9 to obtain a scale-free network topology ($R^2 = 0.85$) and merged highly correlated modules that had an average distance of < 0.65. For the functional association of modules to specific mechanism of action (MOA), we obtained the MOA terms for each drug (from rcellminer) and tested for their overrepresentation in the modules (with a significance threshold of Bonferroni-adjusted p-value < 0.05) where the significant terms were used for the functional annotation of the drug modules.

We proceeded to measure the correlation of the drug modules (or their eigenvalues, acquired through singular value decomposition) to metabolite levels of the cells within the NCI-60 (Ortmayr et al., 2019). An R-script for this network analysis is available as a supplementary file in the online version of Paper II.

3.5 Statistical analysis

For normally distributed data, we employed Student's t-test for the pairwise comparison of treatments. Benjamini-Hochberg adjustment for multiple comparisons was performed when appropriate (Benjamini & Hochberg, 1995). Analysis of variance (ANOVA) was used to compare data from three or more treatments or the simultaneous evaluation of the effect of two grouping variables.

For non-normally distributed data, we used a Mann-Whitney U test to compare two, independent treatments (with a Benjamini-Hochberg adjustment when appropriate). For the comparison of more than two groups, a Kruskal-Wallis test was employed.

All statistical analysis was carried out using the statistical software R (R Core Team, 2020) and MATLAB (Math Works, 1996).

4 Results and discussion

In this chapter, results of each paper are described and discussed separately along with the main conclusions and limitations after which the future perspectives are outlined.

4.1 Paper I: EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition

The aim of this study was to trace isotope-labelled carbon sources throughout the central carbon metabolism in breast cells before and after EMT in D492 and D492M cells, respectively, and to identify the characteristics and vulnerabilities of the post-EMT metabolic phenotype, arguing that EMT is a process that is crucial for metastatic development in cancers.

In a study from 2017, Halldorsson et al. found that the two carbon sources with the highest uptake values in D492 and D492M cells were glucose and glutamine (Halldorsson et al., 2017). In this follow-up study, we sought to investigate and compare the contributions of these two carbon sources specifically to the central carbon metabolic pathways in the two cell lines to further the understanding of EMT-linked metabolic alterations in breast cells. To this end, we grew the cells in media containing three differently labelled carbon sources: (i) 100% ^{13}C-labeled glutamine at position 1, (ii) 100% ^{13}C-labeled glutamine at position 5 and (iii) 100% ^{13}C-labeled glucose at positions 1 and 2. The two differently labelled glutamines allow for the discrimination between reductive and oxidative glutamine metabolism (Gameiro et al., 2013) whereas this specifically labelled glucose has been shown to provide precise estimates of glycolytic activity compared to other differently labelled glucose moieties (Metallo et al., 2009). We found that there was a significant difference in the way the D492 cells use these carbon sources before and after EMT.

4.1.1 Results

First, we found that following EMT, there was an overall decrease in glycolytic activity supported by both (i) a lowered contribution of the 1,2-^{13}C-glucose to lactate inside the cells and (ii) decreased consumption of glucose

and secretion of lactate levels in the cells (**Paper I, Figure 1**). The contribution of glucose to citrate and palmitate was not significantly affected, indicating that the main difference in glucose utilization following EMT is the lowered glycolytic activity which is not translated to the TCA cycle.

Second, we saw that there was a relative increase in the reductive carboxylation (RC) of glutamine-derived alpha-ketoglutarate to citrate and palmitate (**Paper I, Figure 2**). This suggest that there is a difference in the redox potential of the cells, due to a specific reaction in the RC pathway catalysed by isocitrate dehydrogenase that is NADPH-dependent. We confirmed that there was indeed a difference in the redox balance of the cells by measuring the NADPH/NADP ratio in the cells (**Paper I, Figure 3)** which was significantly higher in the D492M cell line. Furthermore, there was a significant decrease in the contribution of glutamine to glutathione and the overall intracellular glutathione levels (**Paper 1, Figure 3**). Glutathione is known for its importance as an antioxidant but also as a drug-neutralizing conjugator in xenobiotic metabolism (Forman et al., 2009; K. Zhang et al., 1998). Because of the significant lowering of its levels following EMT, we hypothesised that this might result in the decreased drug tolerance of D492M cells. To test this hypothesis, we used an integrated network analysis of 214 drug sensitivity profiles of 59 cell lines (from the NCI-60 Human Tumour Cell Line panel (Shoemaker, 2006) and their metabolomic profiles (Ortmayr et al., 2019), which suggested that mTOR-inhibiting drugs would be highly affected by altering the glutathione levels of cells (**Paper I, Figure 5**). To validate the findings from the network analysis, we treated D492 and D492M cells with mTOR inhibitor everolimus and found that indeed, the D492M cells were more sensitive than D492 cells (**Paper** I**, Figure 5C**). To establish whether this difference was observed due to different glutathione levels, we cotreated the cells with buthionine sulphoximine (BSO), a glutathione biosynthesis inhibitor, and everolimus and saw that the sensitivity was significantly increased compared to everolimus treatment alone (**Paper I, Figure 5F**). Importantly, these effects were not observed when the cells were cotreated with BSO and a paclitaxel, a drug targeting microtubule stabilisation.

4.1.2 Conclusions

We traced the contribution of glucose and glutamine to the central carbon metabolism using isotope labelling experiments of normal breast tissue cell lines before and after EMT. Although the paper is highly descriptive, the most significant finding was how the EMT-linked rewiring of glutamine metabolism can affect the sensitivity of cells to mTOR inhibition through diminished

glutathione biosynthesis. To aid with drug selection, we used a novel approach by the integrated network analysis of the drug sensitivity profiles and a metabolomic analysis of the NCI-60 Human Tumour Cell lines (Ortmayr et al., 2019; Shoemaker, 2006). This approach allowed us to identify which drug groups (i.e*.,* drugs that have similar effects on various cell lines) are correlated with a metabolite of interest. Importantly, this approach could also prove useful in future studies focusing on metabolism and drugs, which is why we made all our code available online. The findings from this paper address the first specific aim of this thesis, i.e., to investigate the contribution of glucose and glutamine to central carbon metabolism in the breast EMT cellular model of D492 and D492M cells.

4.1.3 Limitations

The main limitation of this study is the lack of experimental data regarding the contribution of other carbon sources to central carbon metabolism of D492 and D492M cells. For example, there is a substantial concentration of pyruvate (0.5 mM) in the H14 medium which the D492 and D492M cells are grown in. High levels of extracellular pyruvate and lactate provides a large buffer system that significantly slows down the time it takes intracellular pyruvate/lactate to reach isotopic steady-state when tracing carbons from pyruvate, glucose or other upstream carbon sources. This is the main reason why it is typically recommended to remove pyruvate/lactate from the media prior to isotope tracing of glucose (Antoniewicz, 2018). However, we decided not to change the concentrations of pyruvate and lactate in the H14 defined media in order to make our results comparable to decades of research on D492/D492M phenotypic characterization in H14 media. It has been shown that D492 and D492M cells take up carbon sources other than glucose and glutamine in high amounts. In particular, the essential amino acids lysine and threonine, along with conditionally essential amino acid arginine have high uptake values in both cell lines (Halldorsson et al., 2017). The measurement of the incorporation of isotopes from all the high-uptake carbon sources at regular time intervals, e.g*.,* from 1-48 hours, would have allowed us to assess when exactly isotopic steady state is reached and subsequently use the gold standard in fluxomics, ^{13}C-metabolic flux analysis (13C-MFA) (Long & Antoniewicz, 2019) to obtain a more detailed metabolic characterisation of the cells.

4.2 Paper II: Argininosuccinate lyase is a Metabolic Vulnerability in Breast Development and Cancer

The aim of this study was to evaluate the differences in metabolic predictions by GSMMs constrained with different types of omics data. The chosen GSMM was a previous metabolic reconstruction of breast metabolism (Halldorsson et al., 2017), which we termed iBreast2886. We constrained intracellular and extracellular reactions according to (i) microarray data, (ii) RNA-seq data, (iii) proteomic data and (iv) only measurements of extracellular metabolites over time.

4.2.1 Results

The experimental data were obtained from the D492 and D492M breast cell lines. We evaluated the accuracy of metabolic predictions made by the models by comparing them to ^{13}C-labeled glutamine tracing data. To this end, we utilised a metabolic route activity measure which we specifically developed for this task. The GSMMs constrained with proteomic data showed the highest consistency with the ^{13}C tracing data (**Paper II, Figure 3**), which led us to further assess the predictions made by these models in vitro.

As we were dealing with two proteomic GSMMs representing the metabolic phenotypes before and after EMT, we distinguished the GSMMs as EPI and MES, for epithelial and mesenchymal, respectively. The GSMMs share topological properties but had different reaction constraints and uptake/secretion rates. Therefore, we were interested in how reaction bounds of EPI could be minimally altered so that it could take on a flux phenotype similar to MES, and what reactions absolutely require their bounds to be changed for these purposes. To this end, we employed an algorithm developed in a previous study from our laboratory (Choudhary et al., 2016) and found that the reactions that require their bounds to be altered in EPI to be able to have a MES-like flux phenotype were significantly enriched with cholesterol metabolism-related terms (**Paper II, Figure 4**), suggesting that this pathway requires substantial alteration in activity for EMT to take place. To validate these predictions, we treated the D492 and D492M cells with lovastatin, a drug that inhibits cholesterol biosynthesis, and found that D492 cells were more sensitive to the drug in terms of both relative cholesterol abundance reduction and survival, confirming that there is indeed a difference in cholesterol metabolism following EMT in breast.

In addition to predicting pathway activity, GSMMs have proven extremely useful in predicting metabolic vulnerabilities in both bacteria (Monk et al.,

2017) and multicellular organisms (Baloni et al., 2021; Larsson et al., 2020; Pacheco et al., 2019) through simulation of a series of single (or more) gene knockouts and assessing if biomass production is decreased. We performed single gene knockout analysis using the EPI and MES models to identify which genes were predicted to be exclusively essential for biomass production in MES. We found that out of 2140 genes in the models, only 10 were solely essential for biomass production in MES and not EPI. In order to find a single target to pursue in vitro, we ranked the genes based on the concordance index (Harrell's C-index (Harrell et al., 1982)) of survival models based on their expression predicting the overall survival of claudin-low breast cancer patients. This index is a goodness-of-fit measure that quantifies concordance between a risk score and time-to-event, where a value of 1 indicates perfect concordance between survival time and risk score of the patients, but a value of 0.5 indicates completely mixed association of the two. Claudin-low breast tumours have a high expression of EMT markers, and we argued that the survival of patients with this cancer type would be connected to expression of key EMT genes. The top candidate turned out to be argininosuccinate lyase (*ASL*), which we knocked down in vitro using siRNA, resulting in the significant inhibition of survival only in D492M cells, but not in D492 cells, confirming the models' predictions (**Paper II, Figure 5**). Other high-ranking genes (guanylate kinase 1; *GUK1*) were also screened in the same manner and yielded similar results.

After the successful in vitro validation of the GSMM predictions, we were interested in whether the applicability of iBreast2886 could be translated into a clinical setting. Thus, we proceeded to constraining the iBreast2886 model with proteomic data from breast cancer patients (Tang et al., 2018) to identify potential metabolic vulnerabilities of breast cancer on a subtype-specific level. We constructed 65 personalised breast cancer GSMMs, of which 33 were from patients with oestrogen receptor negative (ER^-) tumours and 32 from patients with oestrogen receptor positive ER^+ tumours. After running the single gene knockout simulation for all the GSMMs, we saw that the patients with ER^- breast tumours had a significant over-representation ($p = 0.0419$) of a single essential gene: *ASL*. Multivariable Cox-proportional-hazard models showed that indeed, proteomic levels of *ASL* had moderate effects on the survival of patients with ER^- breast tumours but not on the survival of patients with ER^+ breast tumours (**Paper II, Table 1**).

4.2.2 Conclusions

We focused on the comparison of GSMMs constrained with different omics data. To validate which GSMMs were most accurate in terms of metabolic activity predictions, we compared the results of random sampling to ^{13}C-labelled glutamine data. We found that GSMMs constrained with proteomic data had the highest similarity to the ^{13}C glutamine data, which was not unexpected as it should directly reflect the abundance of metabolic enzymes in the cells. This finding answers the second specific aim of this thesis, the assessment of which omics data integration results in the most accurate metabolic activity predictions. The third specific aim of this thesis was also addressed by the integration of proteomic data from ER^- breast tumours with iBreast2886 to propose metabolic vulnerabilities that had significant association with the survival of ER^- patients only. The main metabolic vulnerability predicted (and validated) for both the post-EMT cell line D492M and ER^- breast tumours was ASL, an enzyme that is vital for *de novo* arginine synthesis, suggesting that it becomes particularly important in these cell/tissue types. Overall, this study presented how context specific GSMMs can be used as tools for the integration of both cellular and clinical omics data to propose biologically relevant targets for therapeutics or prognostic evaluation.

4.2.3 Limitations

The main biological limitation of the study is its lack of detailed characterisation of ASL and why exactly it is essential for post-EMT D492M cells only. To further investigate the role of ASL, the uptake of arginine and its contribution of carbon and/or nitrogen atoms to the urea cycle and other metabolic pathways could be traced and compared in D492 and D492M cells. Moreover, the cells could be deprived of extracellular arginine during growth, and their survival potential subsequently measured to assess whether either cell line displays arginine auxotrophy. Cancer cells have been shown to become reliant on arginine uptake following epigenetic silencing of the ASS1/ASL node. This was assumed to occur to increase the levels of intracellular aspartate and/or polyamines which are necessary for proliferating cells (Pavlova & Thompson, 2016). Arginine is also a precursor for proline, which has repeatedly been linked to EMT induction, invasion and metastasis (Burke et al., 2020; Elia et al., 2017). In fact, we showed that proline metabolism was significantly different in D492 and D492M cells based on isotope-tracing of glutamine.

There were additional steps we could have added to the genome-scale metabolic modelling procedure to ensure more accurate flux predictions from the constraint-based analysis. First of all, the exact composition of the biomass of D492 and D492M cells was not determined nor used for the biomass function in iBreast2886. Instead, the generic biomass function within Recon 2 was used. This biomass function was used when performing FBA, so altering its composition and/or coefficients might alter the intracellular flux profiles of the cells, as has been shown in previous studies of both prokaryotes and eukaryotes (Dikicioglu et al., 2015; Lakshmanan et al., 2019; Pramanik & Keasling, 1998). However, the results from gene knockout simulations have been shown to be quite robust to alterations of the biomass function formulation in both bacteria and human cell lines (Cankorur-Cetinkaya et al., 2017; Opdam et al., 2017), suggesting that the main cell-specific candidates would not be affected by this procedure. Moreover, a comprehensive study of Chinese hamster ovary cells revealed that detailed exchange rates of metabolites, rather than the exact biomass composition, dominate the accuracy of GSMM predictions of growth rates (Széliová et al., 2020). Another way of constraining the solution space even further would be to use the ^{13}C isotope tracing data (e.g., from **Paper I**) as constraints for the model using a similar methodology as previously described (García Martín et al., 2015). Finally, a high-throughput RNA-interference screening could be performed to identify all metabolic genes that are essential for the survival/growth of D492 and D492M cells. By systematically comparing the RNA interference results to the results from a gene knockout simulation of the D492 and D492M constraint-based models, the model parameters could be improved, gaps could be filled, and redundant reactions removed (Thiele & Palsson, 2010).

One of the main concerns about iBreast2886 is that the data used for its construction originate from two cell lines, D492 and D492M, which raises questions about its generality. To further generalise the iBreast2886 GSMM, the transcriptome, proteome, extracellular uptake/secretion rates and RNA interference screening results from other normal breast tissue cell lines (preferably with an EMT cell line counterpart like D492M, e.g., HMLE (Tam et al., 2013) and PMC42 (Ackland et al., 2003)) could be measured and incorporated. In this case, the solution space of the resulting general breast genome-scale metabolic reconstruction should be more generalisable, making it more suitable for a broader application in breast metabolism research, including clinical applications.

4.3 Paper III: Glutamine-fructose-6-phosphate transaminase 2 (GFPT2) is upregulated in breast epithelial-mesenchymal transition and responds to oxidative stress

The aim of this study was to identify specifically the changes in proteomic levels of enzymes in the D492 cell line following EMT (using the aforementioned D492-D492M cell culture model) and induced tumourigenecity by HER2 overexpression. For the latter part, a previously generated HER2-overexpressing D492 cell line (from now on referred to as D492-HER2) was used. This cell line has been shown to be tumourigenic and described as having a partial mesenchymal phenotype, thus having undergone some degree of EMT (Ingthorsson, Andersen, et al., 2016). The comparative analysis of these cell lines was expected to increase the understanding of the role of EMT in breast tumourigenesis.

4.3.1 Results

In order to distinguish the cell lines that have undergone EMT from the epithelial cell line D492, the proteomic profiles of the cells were compared. Compared to D492 cells, the D492M and D492-HER2 cell lines had significantly different expression of enzymes involved with glycan metabolism (**Paper III, Figures 3A-C**). Enzymes that are closely associated with the metabolism of glycan precursors were found to have more than twofold difference in proteomic levels across the three cell lines. These were PYGB, PGM3, UGDH, PGM2L1, GALNT7, GFPT2, and GALE, of which the highest difference in both RNA and protein levels was in GFPT2 (**Paper III, Figure 4**). GFPT2, or Glutamine-Fructose-6-phosphate Transaminase 2, is a rate-limiting enzyme in the hexosamine biosynthetic pathway which regulates the production of precursors for O-GlcNAcylation. Due to its clear upregulation in both EMT-derived D492M and D492-HER2 cells, we hypothesised that manipulating its levels would affect the mesenchymal status of the cells.

Inhibiting the expression of GFPT2 using siRNA-mediated knockdown resulted in the shift from a mesenchymal state to more epithelial one as reflected in the downregulation of known EMT marker vimentin (**Paper III, Figures 4E and F**). Furthermore, there was a significant inhibition of growth of D492M and D492-HER2 cells following GFPT2 downregulation, which was not observed in D492 cells (**Paper III, Figures 4J-L**) along with decreased invasive and migratory capabilities of the cells (**Paper III, Figures 4M-O**). To validate the generality of these results, we investigated the expression of GFPT2 in various breast cancer cell lines two different databases, CCLE and

HMS LINCS. We saw that increased exression of GFPT2 was a trait commonly observed in claudin-low cell lines, rather than HER2-enriched cell lines (**Paper III, Figure 5**).

To confirm that there was indeed an EMT-linked difference in the production rate of glycan precursors, we performed targeted measurements of the intracellular metabolome of the cells in addition to performing an isotope tracer analysis after culturing the cells with ^{13}C-labelled carbon sources. It was clear that among different glycan precursors, there was only a significant upregulation in the intracellular levels of GFPT2 product UDP-GlcNAc following EMT (**Paper III, Figure 6B**). Additionally, there was an increased incorporation of both glucose and glutamine to that metabolite in both D492M and D492-HER2 cells compared to D492 cells (**Paper III, Figure 6C**). By downregulating GFPT2 using siRNA, there was a significant reduction of intracellular UDP-GlcNAc levels in the cells, (**Paper III, Figure 6D**), along with a decreased in the levels of glutamate and cystathionine (**Paper III, Figures 6E and F**). Glutamate and cystathionine are both precursors for glutathione (GSH) generation, so we hypothesised that the GFPT2 enzyme was linked to the redox status of the cells. Indeed, a gene-metabolite correlation analysis of the cells from the NCI-60 Human Tumour Cell line panel showed that there was an inverse relationship between GSH abundance and *GFPT2* expression (**Paper III, Figure 7A**). Accordingly, we found that by treating claudin-low cell line MDA-MB-231 cell line with GSH resulted in lowered expression of *GFPT2*, suggesting that the enzyme reacts to GSH (**Paper III, Figure 7E**). Both the D492M and D492-HER2 cell lines have lower levels of GSH than D492 cells (**Paper III, Figure 7F**) which is in accordance with their higher GFPT2 protein abundance. In addition to being a part of GSH metabolism, cystathionine is a precursor for the synthesis of the antioxidant H_2S through the transsulfuration pathway. We found that by downregulating *GFPT2* expression, there was a significantly reduced expression of Sulfide Quinone Oxidoreductase (*SQOR*) in D492, D492M, D492-HER2 and MDA-MB-231 cell lines (**Paper III, Figures 7G-N**), which is a gene coding for an enzyme that uses H_2S to regenerate ubiquinol in the mitochondrial membrane. This further suggests that there is a functional relationship between GFPT2 and the oxidative stress response. Finally, we found that GFPT2 is regulated by both EGF and insulin, and inhibited by GSK3-β, supported by growth factor manipulation studies, phosphoproteomic analysis and siRNA knockdown experiments (**Paper III, Figure 8**).

4.3.2 Conclusions

In this study, we have described GFPT2 as an important enzyme in EMT and the oxidative stress response of breast cells along with being a potential biomarker for claudin-low breast cancer. In general, the study demonstrates how interrogating omics data in an ordered fashion may lead to the identification of a single protein that is functionally relevant in the context of EMT and tumourigenesis.

4.3.3 Limitations

A large part of the results were presented only in the expanded D492 cell culture model (i.e., D492, D492M, D492-HER2) which provides a limited vision of normal and cancerous breast tissue. Even within the D492 cell culture model, there is ectopic overexpression of HER2 (ErbB2) which is not „naturally“ induced EMT like in the case of D492-to-D492M.

In the study, a few avenues for further research are speculated: (i) how GFPT2 affects cystathionine, (ii) the link between GFPT2, oxidative stress and H_2S/SQOR and (iii) the generality of the effects of insulin, EGF and GSK3-β on GFPT2. The first could be addressed by expanding the isotope tracing methodology to trace metabolites that are involved with the transsulfuration pathway (e.g., 3-^{13}C-L-serine (Zhu et al., 2019)) before and after genetic manipulation of GFPT2. The second avenue could be addressed by performing more oxidative stress response experiments where ROS is introduced to a cell (preferably multiple cell lines) and H_2S synthesis is monitored over time using specific assays. This is done before and after GFPT2 is genetically manipulated. Finally, the third avenue could be addressed by simply performing the same growth factor manipulation studies and siRNA mediated inhibition of GFPT2 expression in other cell lines.

4.4 Future perspective

The main limitations of the papers have been discussed along with potential ways to address them. However, what generally remains to be discussed is the future research and the opportunities and pitfalls that are on the horizon.

A 13C-MFA of the data from stable isotope-tracing of multiple carbon sources in D492 and D492M would undeniably result in a quantitatively accurate description of their central carbon metabolism. However, its clinical applicability has been questioned due to the unavailability of large-scale fluxomic data for integrated, computational modelling of context-specific metabolism (Dai & Locasale, 2017), in which the usage of genome-scale

metabolic modelling has proven useful. Manually curated context-specfic GSMMs provide mechanistic insights into cellular metabolism by integrating genomic, transcriptomic, proteomic and metabolomic data to produce fluxomic profiles that allow functional characterisation of genes, as opposed to their abstract characterisation based on genome annotation alone. The generation of GSMM fluxomic data is also both relatively cheap and fast compared to other types of omic data and is therefore an excellent additional numerical data in the multi-omic analysis of organisms (Zampieri et al., 2019). In recent years, various studies have shown that the inclusion of fluxomic data (or gene knockout simulation results) in machine learning algorithms has increased the accuracy of phenotype predictions compared to transcriptomic data alone in both prokaryotes and eukaryotes (Culley et al., 2020; Guo et al., 2017; Yaneske & Angione, 2017; Zampieri et al., 2017). These findings indicate that the therapeutics and prognostics of breast cancer could be improved by leveraging machine learning models with omics-based fluxomic data.

In this thesis, multiple dimensions of D492 and D492M cells have been measured and/or estimated. This includes the transcriptome, proteome, metabolome and fluxome. However, one dimension remains to be assessed – the epigenome. It is important to acknowledge that there is a link between the epigenetic landscape, post-translational modifications and metabolism, due to certain metabolites being either cofactors for epigenetic regulators (e.g*.,* alpha-ketoglutarate for TET enzymes) or direct conjugators of DNA, histones and/or other proteins (e.g*.,* methyl-, and acetyl groups) (Wong et al., 2017). Even though this relationship was beyond the scope of this thesis, its investigation is imperative for the overall understanding of the functional relationship between metabolism and transcriptional regulation. Experiments should be conducted where the production of epigenetic metabolic cofactors is influenced, followed by assessment of the epigenome (e.g*.,* bisulfite sequencing for DNA methylation and MS^2 for post-translational modification of proteins), transcriptome and proteome.

An exciting avenue for further research is the plasticity of EMT, where the varying phenotypes in the EMT spectrum are investigated as a dynamic process. In order to investigate this, however, one would first require a cellular model of the EMT plasticity, where it can be induced in small steps. That would require more knowledge about the activity of TFs and signalling pathways along the way, and how the process is regulated. The dynamics of EMT is certainly a hot topic, and it has been proposed that systems biology

approaches will help elucidating the connection of metabolism and EMT plasticity in the near future (Jia et al., 2021).

In summary, we have contributed to the expansion of the existing knowledge of (i) EMT metabolism and its functional relationship to drug resistance. (ii) GSMM and omics integration for metabolic activity inference and vulnerability detection in both normal tissue and cancer and (iii) the regulation of glycan metabolism in EMT of normal breast tissue and breast cancer. In particular, this work has highlighted how systems biological approaches can be used to guide research in molecular biology and thus combine the holistic and reductionist ideologies for the overall advancement of biological sciences.

References

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. *Lab. Investig.*, *83*(3), 435–448. https://doi.org/10.1097/01.LAB.0000059927.97515.FD

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. & Nielsen, J. (2012). Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT. *PLoS Comput. Biol.*, *8*(5). https://doi.org/10.1371/JOURNAL.PCBI.1002518

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M. & Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.*, *10*(3), 721. https://doi.org/10.1002/msb.145122

Albert, R. & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, *74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47

Alonso, A., Julià, A., Beltran, A., Vinaixa, M., Díaz, M., Ibañez, L., Correig, X. & Marsal, S. (2011). AStream: An R package for annotating LC/MS metabolomic data. *Bioinformatics*, *27*(9), 1339–1340. https://doi.org/10.1093/bioinformatics/btr138

Antoniewicz, M. R. (2018). A guide to 13C metabolic flux analysis for the cancer biologist. *Exp. Mol. Med. 2018 504*, *50*(4), 1–13. https://doi.org/10.1038/s12276-018-0060-y

Arkin, A., Shen, P. & Ross, J. (1997). A test case of correlation metric construction of a reaction pathway from measurements. *Science (80-.).*, *277*(5330), 1275–1279. https://doi.org/10.1126/SCIENCE.277.5330.1275

Aron, A. T., Gentry, E. C., McPhail, K. L., Nothias, L. F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J. M., Sikora, N., Vargas, F., van der Hooft, J. J. J., Ernst, M., Kang, K. Bin, Aceves, C. M., Caraballo-Rodríguez, A. M., Koester, I., Weldon, K. C., Bertrand, S., Roullier, C., … Dorrestein, P. C. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.*, *15*(6), 1954–1991. https://doi.org/10.1038/s41596-020-0317-5

Asgari, Y., Khosravi, P., Zabihinpour, Z. & Habibi, M. (2018). Exploring candidate biomarkers for lung and prostate cancers using gene expression and flux variability analysis. *Integr. Biol. (United Kingdom)*, *10*(2), 113–120. https://doi.org/10.1039/c7ib00135e

Baloni, P., Dinalankara, W., Earls, J. C., Knijnenburg, T. A., Geman, D., Marchionni, L. & Price, N. D. (2021). Identifying personalized metabolic signatures in breast cancer. *Metabolites*, *11*(1), 1–14. https://doi.org/10.3390/metabo11010020

Barabási, A. L., Gulbahce, N. & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.*, *12*(1), 56–68. https://doi.org/10.1038/nrg2918

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bhattacharya, D. & Scimè, A. (2019). Metabolic Regulation of Epithelial to Mesenchymal Transition: Implications for Endocrine Cancer. *Front. Endocrinol. (Lausanne).*, *0*, 773. https://doi.org/10.3389/FENDO.2019.00773

Bhowmik, S. K., Ramirez-Peña, E., Arnold, J. M., Putluri, V., Sphyris, N., Michailidis, G., Putluri, N., Ambs, S., Sreekumar, A. & Mani, S. A. (2015). EMT-induced metabolite signature identifies poor clinical outcome. *Oncotarget*, *6*(40), 42651–42660. https://doi.org/10.18632/oncotarget.4765

Bidkhori, G., Benfeitas, R., Elmas, E., Kararoudi, M. N., Arif, M., Uhlen, M., Nielsen, J. & Mardinoglu, A. (2018). Metabolic network-based identification and prioritization of anticancer targets based on expression data in hepatocellular carcinoma. *Front. Physiol.*, *9*(JUL). https://doi.org/10.3389/fphys.2018.00916

Bidkhori, G., Benfeitas, R., Klevstig, M., Zhang, C., Nielsen, J., Uhlen, M., Boren, J. & Mardinoglu, A. (2018). Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(50), E11874–E11883. https://doi.org/10.1073/pnas.1807305115

Björnson, E., Mukhopadhyay, B., Asplund, A., Pristovsek, N., Cinar, R., Romeo, S., Uhlen, M., Kunos, G., Nielsen, J. & Mardinoglu, A. (2015). Stratification of Hepatocellular Carcinoma Patients Based on Acetate Utilization. *Cell Rep.*, *13*(9), 2014–2026. https://doi.org/10.1016/j.celrep.2015.10.045

Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Briem, E., Budkova, Z., Sigurdardottir, A. K., Hilmarsdottir, B., Kricker, J., Timp, W., Magnusson, M. K., Traustadottir, G. A. & Gudjonsson, T. (2019). MiR-203a is differentially expressed during branching morphogenesis and EMT in breast progenitor cells and is a repressor of peroxidasin. *Mech. Dev.*, *155*, 34–47. https://doi.org/10.1016/J.MOD.2018.11.002

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. *J. Mammary Gland Biol. Neoplasia*, *24*(2), 139–147. https://doi.org/10.1007/s10911-018-09424-w

Broido, A. D. & Clauset, A. (2019). Scale-free networks are rare. *Nat. Commun. 2019 101*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-08746-5

Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlic, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., … Palsson, B. O. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.*, *36*(3), 272–281. https://doi.org/10.1038/nbt.4072

Buescher, J. M., Antoniewicz, M. R., Boros, L. G., Burgess, S. C., Brunengraber, H., Clish, C. B., DeBerardinis, R. J., Feron, O., Frezza, C., Ghesquiere, B., Gottlieb, E., Hiller, K., Jones, R. G., Kamphorst, J. J., Kibbey, R. G., Kimmelman, A. C., Locasale, J. W., Lunt, S. Y., Maddocks, O. D. K., … Fendt, S. M. (2015). A roadmap for interpreting 13C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.*, *34*, 189–201. https://doi.org/10.1016/j.copbio.2015.02.003

Burke, L., Guterman, I., Palacios Gallego, R., Britton, R. G., Burschowsky, D., Tufarelli, C. & Rufini, A. (2020). The Janus-like role of proline metabolism in cancer. *Cell Death Discov. 2020 61*, *6*(1), 1–17. https://doi.org/10.1038/s41420-020-00341-8

Cankorur-Cetinkaya, A., Dikicioglu, D. & Oliver, S. G. (2017). Metabolic modeling to identify engineering targets for Komagataella phaffii: The effect of biomass composition on gene target identification. *Biotechnol. Bioeng.*, *114*(11), 2605–2615. https://doi.org/10.1002/bit.26380

Chan, S. Y. & Loscalzo, J. (2012). The emerging paradigm of network medicine in the study of human disease. *Circ. Res.*, *111*(3), 359–374. https://doi.org/10.1161/CIRCRESAHA.111.258541

Choudhary, K. S., Rohatgi, N., Halldorsson, S., Briem, E., Gudjonsson, T., Gudmundsson, S. & Rolfsson, O. (2016). EGFR Signal-Network Reconstruction Demonstrates Metabolic Crosstalk in EMT. *PLoS Comput. Biol.*, *12*(6), e1004924. https://doi.org/10.1371/journal.pcbi.1004924

Clasquin, M. F., Melamud, E. & Rabinowitz, J. D. (2012). LC-MS data processing with MAVEN: A metabolomic analysis and visualization engine. *Curr. Protoc. Bioinforma.*, *Chapter 14*(SUPPL.37). https://doi.org/10.1002/0471250953.bi1411s37

Culley, C., Vijayakumar, S., Zampieri, G. & Angione, C. (2020). A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. *Proc. Natl. Acad. Sci.*, *117*(31), 18869–18879. https://doi.org/10.1073/PNAS.2002959117

Dai, Z. & Locasale, J. W. (2017). Understanding metabolism with flux analysis: From theory to application. *Metab. Eng.*, *43*, 94–102. https://doi.org/10.1016/J.YMBEN.2016.09.005

Dai, Z., Yang, S., Xu, L., Hu, H., Liao, K., Wang, J., Wang, Q., Gao, S., Li, B. & Lai, L. (2019). Identification of Cancer–associated metabolic vulnerabilities by modeling multi-objective optimality in metabolism. *Cell Commun. Signal. 2019 171*, *17*(1), 1–15. https://doi.org/10.1186/S12964-019-0439-Y

De Souza Palma, C., Grassi, M. L., Thomé, C. H., Ferreira, G. A., Albuquerque, D., Pinto, M. T., Melo, F. U. F., 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. https://doi.org/10.1074/MCP.M115.052910

Dikicioglu, D., Kırdar, B. & Oliver, S. G. (2015). Biomass composition: the “elephant in the room” of metabolic modelling. *Metabolomics*, *11*(6), 1690–1701. https://doi.org/10.1007/s11306-015-0819-2

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R. & Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U. S. A.*, *104*(6), 1777. https://doi.org/10.1073/PNAS.0610772104

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. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.*, *6*(7), 1060–1083. https://doi.org/10.1038/nprot.2011.335

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J. & Viant, M. R. (2013). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, *9*(SUPPL.1), 44–66. https://doi.org/10.1007/s11306-012-0434-4

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. https://doi.org/10.1093/bioinformatics/18.suppl_1.S105

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. *Int J Biochem Cell Biol*, *103*, 99–104. https://doi.org/10.1016/j.biocel.2018.08.003

Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., Verfaillie, C., Grünewald, 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. https://doi.org/10.1038/ncomms15267https://www.nature.com/articles/ncomms15267#supplementary-information

Emwas, A. H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Nagana Gowda, G. A., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M. & Wishart, D. S. (2019). NMR spectroscopy for metabolomics research. *Metabolites*, *9*(7). https://doi.org/10.3390/metabo9070123

Fang, N., Yu, S., Ronis, M. J. J. & Badger, T. M. (2015). Matrix effects break the LC behavior rule for analytes in LC-MS/MS analysis of biological samples. *Exp. Biol. Med.*, *240*(4), 488–497. https://doi.org/10.1177/1535370214554545

Fang, X., Lloyd, C. J. & Palsson, B. O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol. 2020 1812*, *18*(12), 731–743. https://doi.org/10.1038/s41579-020-00440-4

Farhadian, M., Rafat, S. A., Panahi, B. & Mayack, C. (2021). Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. *Sci. Reports 2021 111*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-81888-z

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1990). Electrospray ionization–principles and practice. *Mass Spectrom. Rev.*, *9*(1), 37–70. https://doi.org/10.1002/mas.1280090103

Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics*, *2*(3), 155–168. https://doi.org/10.1002/cfg.82

Fiehn, Oliver. (2002). Metabolomics — the link between genotypes and phenotypes. *Funct. Genomics*, 155–171. https://doi.org/10.1007/978-94-010-0448-0_11

Forman, H. J., Zhang, H. & Rinna, A. (2009). Glutathione: Overview of its protective roles, measurement, and biosynthesis. *Mol. Aspects Med.*, *30*(1–2), 1–12. https://doi.org/10.1016/j.mam.2008.08.006

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Fruchterman, T. M. J. & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Softw. Pract. Exp.*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Fukushima, A., Kusano, M., Redestig, H., Arita, M. & Saito, K. (2011). Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. *BMC Syst. Biol. 2011 51*, *5*(1), 1–12. https://doi.org/10.1186/1752-0509-5-1

Gameiro, P. A., Yang, J., Metelo, A. M., Pérez-Carro, R., Baker, R., Wang, Z., Arreola, A., Rathmell, W. K., Olumi, A., López-Larrubia, P., Stephanopoulos, G. & Iliopoulos, O. (2013). In vivo HIF-mediated reductive carboxylation is regulated by citrate levels and sensitizes VHL-deficient cells to glutamine deprivation. *Cell Metab.*, *17*(3), 372–385. https://doi.org/10.1016/j.cmet.2013.02.002

García Martín, H., Kumar, V. S., Weaver, D., Ghosh, A., Chubukov, V., Mukhopadhyay, A., Arkin, A. & Keasling, J. D. (2015). A Method to Constrain Genome-Scale Models with 13C Labeling Data. *PLoS Comput. Biol.*, *11*(9), e1004363. https://doi.org/10.1371/journal.pcbi.1004363

García Sánchez, C. E., Vargas García, C. A. & Torres Sáez, R. G. (2012). Predictive potential of flux balance analysis of Saccharomyces cerevisiae using as optimization function combinations of cell compartmental objectives. *PLoS One*, *7*(8), e43006. https://doi.org/10.1371/journal.pone.0043006

Grassi, M. L., Palma, C. de S., Thomé, C. H., Lanfredi, G. P., Poersch, A. & Faça, V. M. (2017). Proteomic analysis of ovarian cancer cells during epithelial-mesenchymal transition (EMT) induced by epidermal growth factor (EGF) reveals mechanisms of cell cycle control. *J. Proteomics*, *151*, 2–11. https://doi.org/10.1016/J.JPROT.2016.06.009

Grassian, A. R., Lin, F., Barrett, R., Liu, Y., Jiang, W., Korpal, M., Astley, H., Gitterman, D., Henley, T., Howes, R., Levell, J., Korn, J. M. & Pagliarini, R. (2012). Isocitrate Dehydrogenase (IDH) Mutations Promote a Reversible ZEB1/MicroRNA (miR)-200-dependent Epithelial-Mesenchymal Transition (EMT). *J. Biol. Chem.*, *287*(50), 42180. https://doi.org/10.1074/JBC.M112.417832

Gudjonsson, T., Adriance, M. C., Sternlicht, M. D., Petersen, O. W. & Bissell, M. J. (2005). Myoepithelial Cells: Their Origin and Function in Breast Morphogenesis and Neoplasia. *J. Mammary Gland Biol. Neoplasia*, *10*(3), 261. https://doi.org/10.1007/S10911-005-9586-4

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 Dev.*, *16*(6), 693–706. https://doi.org/10.1101/gad.952602

Gullberg, J., Jonsson, P., Nordström, A., Sjöström, M. & Moritz, T. (2004). Design of experiments: An efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Biochem.*, *331*(2), 283–295. https://doi.org/10.1016/j.ab.2004.04.037

Guo, W., Xu, Y. & Feng, X. (2017). *DeepMetabolism: A Deep Learning System to Predict Phenotype from Genome Sequencing*.

Gysi, D. M., Do Valle, Í., Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S. D., Patten, J. J., Davey, R. A., Loscalzo, J. & Barabási, A. L. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc. Natl. Acad. Sci. U. S. A.*, *118*(19). https://doi.org/10.1073/pnas.2025581118

Hagan, T., Cortese, M., Rouphael, N., Boudreau, C., Linde, C., Maddur, M. S., Das, J., Wang, H., Guthmiller, J., Zheng, N. Y., Huang, M., Uphadhyay, A. A., Gardinassi, L., Petitdemange, C., McCullough, M. P., Johnson, S. J., Gill, K., Cervasi, B., Zou, J., … Pulendran, B. (2019). Antibiotics-Driven Gut Microbiome Perturbation Alters Immunity to Vaccines in Humans. *Cell*, *178*(6), 1313-1328.e13. https://doi.org/10.1016/J.CELL.2019.08.010

Halldorsson, S., Rohatgi, N., Magnusdottir, M., Choudhary, K. S., Gudjonsson, T., Knutsen, E., Barkovskaya, A., Hilmarsdottir, B., Perander, M., Mælandsmo, G. M., Gudmundsson, S. & Rolfsson, Ó. (2017). Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett.*, *396*, 117–129. https://doi.org/10.1016/j.canlet.2017.03.019

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, *33*(DATABASE ISS.). https://doi.org/10.1093/nar/gki033

Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Haraldsdóttir, H. S., Cousins, B., Thiele, I., Fleming, R. M. . & Vempala, S. (2017). CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, *33*(11), 1741–1743. https://doi.org/10.1093/BIOINFORMATICS/BTX052

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J. & Cardoso, F. (2019). Breast cancer. *Nat. Rev. Dis. Prim. 2019 51*, *5*(1), 1–31. https://doi.org/10.1038/s41572-019-0111-2

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. (1982). Evaluating the Yield of Medical Tests. *JAMA J. Am. Med. Assoc.*, *247*(18), 2543–2546. https://doi.org/10.1001/jama.1982.03320430047030

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., … Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.*, *14*(3), 639–702. https://doi.org/10.1038/s41596-018-0098-2

Hennigs, A., Riedel, F., Gondos, A., Sinn, P., Schirmacher, P., Marmé, F., Jäger, D., Kauczor, H.-U., Stieber, A., Lindel, K., Debus, J., Golatta, M., Schütz, F., Sohn, C., Heil, J. & Schneeweiss, A. (2016). Prognosis of breast cancer molecular subtypes in routine clinical care: A large prospective cohort study. *BMC Cancer*, *16*(1). https://doi.org/10.1186/S12885-016-2766-3

Herzog, C., Hook, D. & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quant. Sci. Stud.*, *1*(1), 387–395. https://doi.org/10.1162/QSS_A_00020

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 ΔNp63 are required for breast epithelial differentiation and branching morphogenesis. *Dev. Biol.*, *403*(2), 150–161. https://doi.org/10.1016/J.YDBIO.2015.05.007

Holliday, D. L. & Speirs, V. (2011). Choosing the right cell line for breast cancer research. *Breast Cancer Res.*, *13*(4). https://doi.org/10.1186/bcr2889

Hopkinson, B. M., Klitgaard, M. C., Petersen, O. W., Villadsen, R., Rønnov-Jessen, L. & Kim, J. (2017). Establishment of a normal-derived estrogen receptor-positive cell line comparable to the prevailing human breast cancer subtype. *Oncotarget*, *8*(6), 10580–10593. https://doi.org/10.18632/oncotarget.14554

Ingthorsson, 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. https://doi.org/10.1038/onc.2015.489

Ingthorsson, S., Briem, E., Bergthorsson, J. T. & Gudjonsson, T. (2016). Epithelial Plasticity During Human Breast Morphogenesis and Cancer Progression. *J. Mammary Gland Biol. Neoplasia*, *21*(3–4), 139–148. https://doi.org/10.1007/s10911-016-9366-3

Jahagirdar, S., Suarez-Diez, M. & Saccenti, E. (2019). Simulation and Reconstruction of Metabolite-Metabolite Association Networks Using a Metabolic Dynamic Model and Correlation Based Algorithms. *J. Proteome Res.*, *18*(3), 1099–1113. https://doi.org/10.1021/acs.jproteome.8b00781

Jeong, H., Tombor, B., Albert, R., Oltval, Z. N. & Barabásl, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654. https://doi.org/10.1038/35036627

Jia, D., Park, J. H., Kaur, H., Jung, K. H., Yang, S., Tripathi, S., Galbraith, M., Deng, Y., Jolly, M. K., Kaipparettu, B. A., Onuchic, J. N. & Levine, H. (2021). Towards decoding the coupled decision-making of metabolism and epithelial-to-mesenchymal transition in cancer. *Br. J. Cancer 2021 12412*, *124*(12), 1902–1911. https://doi.org/10.1038/s41416-021-01385-y

Johnson, A., Alberts, B., Bray, D., Hopkin, K. & Raff, M. (2009). *Essential Cell Biology*. Garland Science.

Johnstone, I. M. & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, *367*(1906), 4237–4253. https://doi.org/10.1098/rsta.2009.0159

Jolly, M. K., Boareto, M., Huang, B., Jia, D., Lu, M., Ben-Jacob, E., Onuchic, J. N. & Levine, H. (2015). Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Front. Oncol.*, *0*(JUN), 155. https://doi.org/10.3389/FONC.2015.00155

Jolly, M. K., Tripathi, S. C., Jia, D., Mooney, S. M., Celiktas, M., Hanash, S. M., Mani, S. A., Pienta, K. J., Ben-Jacob, E. & Levine, H. (2016). Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget*, *7*(19), 27067. https://doi.org/10.18632/ONCOTARGET.8166

Kalluri, R. & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J Clin Invest*, *119*(6), 1420–1428. https://doi.org/10.1172/jci39104

Katajamaa, M. & Orešič, M. (2007). Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A*, *1158*(1–2), 318–328. https://doi.org/10.1016/j.chroma.2007.04.021

Konrat, R., Papp, H., Szijártó, V., Gesell, T., Nagy, G., Madai, M., Zeghbib, S., Kuczmog, A., Lanszki, Z., Helyes, Z., Kemenesi, G., Jakab, F. & Nagy, E. (2020). The Anti-histamine Azelastine, Identified by Computational Drug Repurposing, Inhibits SARS-CoV-2 Infection in Reconstituted Human Nasal Tissue In Vitro. *BioRxiv*, 2020.09.15.296228. https://doi.org/10.1101/2020.09.15.296228

Korthauer, K., Kimes, P. K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., Shukla, C., Alm, E. J. & Hicks, S. C. (2019). A practical guide to methods controlling false discoveries in computational biology. *Genome Biol. 2019 201*, *20*(1), 1–21. https://doi.org/10.1186/S13059-019-1716-1

Koschützki, D. & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul. Syst. Bio.*, *2008*(2), 193–201. https://doi.org/10.4137/grsb.s702

Kostiainen, R. & Kauppila, T. J. (2009). Effect of eluent on the ionization process in liquid chromatography–mass spectrometry. *J. Chromatogr. A*, *1216*(4), 685–699. https://doi.org/10.1016/J.CHROMA.2008.08.095

Kotronoulas, A., Jónasdóttir, H. S., Sigurðardóttir, R. S., Halldórsson, S., Haraldsson, G. G. & Rolfsson, Ó. (2020a). Wound healing grafts: Omega-3 fatty acid lipid content differentiates the lipid profiles of acellular Atlantic cod skin from traditional dermal substitutes. *J. Tissue Eng. Regen. Med.*, *14*(3), 441–451. https://doi.org/10.1002/TERM.3005

Kotronoulas, A., Jónasdóttir, H. S., Sigurðardóttir, R. S., Halldórsson, S., Haraldsson, G. G. & Rolfsson, Ó. (2020b). Wound healing grafts: Omega-3 fatty acid lipid content differentiates the lipid profiles of acellular Atlantic cod skin from traditional dermal substitutes. *J. Tissue Eng. Regen. Med.*, *14*(3), 441–451. https://doi.org/10.1002/term.3005

Koutrouli, M., Karatzas, E., Paez-Espino, D. & Pavlopoulos, G. A. (2020). A Guide to Conquer the Biological Network Era Using Graph Theory. *Front. Bioeng. Biotechnol.*, *8*, 34. https://doi.org/10.3389/fbioe.2020.00034

Kruve, A. & Kaupmees, K. (2017). Adduct Formation in ESI/MS by Mobile Phase Additives. *J. Am. Soc. Mass Spectrom.*, *28*(5), 887–894. https://doi.org/10.1007/s13361-017-1626-y

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.*, *84*(1), 283–289. https://doi.org/10.1021/ac202450g

Lakshmanan, M., Long, S., Ang, K. S., Lewis, N. & Lee, D.-Y. (2019). On the impact of biomass composition in constraint-based flux analysis. *BioRxiv*, 652040. https://doi.org/10.1101/652040

Lamouille, S., Xu, J. & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.*, *15*(3), 178–196. https://doi.org/10.1038/nrm3758

Langfelder, P. & Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.*, *1*(1), 54. https://doi.org/10.1186/1752-0509-1-54

Langfelder, P. & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Larsson, I., Uhlén, M., Zhang, C. & Mardinoglu, A. (2020). Genome-Scale Metabolic Modeling of Glioblastoma Reveals Promising Targets for Drug Development. *Front. Genet.*, *11*, 381. https://doi.org/10.3389/fgene.2020.00381

Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., Jones, D. P. & Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.*, *9*(7), 1003123. https://doi.org/10.1371/journal.pcbi.1003123

Li, X., Zhang, X., Ye, L., Kang, Z., Jia, D., Yang, L. & Zhang, B. (2019). LC-MS-Based Metabolomic Approach Revealed the Significantly Different Metabolic Profiles of Five Commercial Truffle Species. *Front. Microbiol.*, *10*, 2227. https://doi.org/10.3389/fmicb.2019.02227

Liu, X. & Locasale, J. W. (2017). Metabolomics: A Primer. *Trends Biochem. Sci.*, *42*(4), 274–284. https://doi.org/10.1016/j.tibs.2017.01.004

Lomana, A. L. G. De, Beg, Q. K., Fabritiis, G. De & Villà-Freixa, J. (2010). Statistical Analysis of Global Connectivity and Activity Distributions in Cellular Networks. *Https://Home.Liebertpub.Com/Cmb*, *17*(7), 869–878. https://doi.org/10.1089/CMB.2008.0240

Long, C. P. & Antoniewicz, M. R. (2019). High-resolution 13C metabolic flux analysis. *Nat. Protoc.*, *14*(10), 2856–2877. https://doi.org/10.1038/s41596-019-0204-0

Loscalzo, J. & Barabasi, A. L. (2011). Systems biology and the future of medicine. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, *3*(6), 619–627. https://doi.org/10.1002/wsbm.144

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.*, *110*(45), 18144–18149. https://doi.org/10.1073/PNAS.1318192110

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., Choi, D., Coté, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., … Calderwood, M. A. (2020). A reference map of the human binary protein interactome. *Nature*, *580*(7803), 402–408. https://doi.org/10.1038/s41586-020-2188-x

Luna, A., Rajapakse, V. N., Sousa, F. G., Gao, J., Schultz, N., Varma, S., Reinhold, W., Sander, C. & Pommier, Y. (2016). Rcellminer: Exploring molecular profiles and drug response of the NCI-60 cell lines in R. *Bioinformatics*, *32*(8), 1272–1274. https://doi.org/10.1093/bioinformatics/btv701

Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O. & Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.*, *3*, 135. https://doi.org/10.1038/msb4100177

Machado, D. & Herrgård, M. (2014). Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLOS Comput. Biol.*, *10*(4), e1003580. https://doi.org/10.1371/JOURNAL.PCBI.1003580

Mardinoglu, A., Bjornson, E., Zhang, C., Klevstig, M., Söderlund, S., Ståhlman, M., Adiels, M., Hakkarainen, A., Lundbom, N., Kilicarslan, M., Hallström, B. M., Lundbom, J., Vergès, B., Barrett, P. H. R., Watts, G. F., Serlie, M. J., Nielsen, J., Uhlén, M., Smith, U., … Boren, J. (2017). Personal model-assisted identification of NAD+ and glutathione metabolism as intervention target in NAFLD. *Mol. Syst. Biol.*, *13*(3), 916. https://doi.org/10.15252/MSB.20167422

Marić, I., Oskotsky, T., Kosti, I., Le, B., Wong, R. J., Shaw, G. M., Sirota, M. & Stevenson, D. K. (2021). Decreased Mortality Rate Among COVID-19 Patients Prescribed Statins: Data From Electronic Health Records in the US. *Front. Med.*, *0*, 30. https://doi.org/10.3389/FMED.2021.639804

Marín de Mas, I., Aguilar, E., Zodda, E., Balcells, C., Marin, S., Dallmann, G., Thomson, T. M., Papp, B. & Cascante, M. (2018). Model-driven discovery of long-chain fatty acid metabolic reprogramming in heterogeneous prostate cancer cells. *PLoS Comput. Biol.*, *14*(1). https://doi.org/10.1371/journal.pcbi.1005914

Math Works. (1996). *MATLAB The language of Technical Computing Installation guide for PC*. Natwick : Math Works Inc., 1996.

McGarrity, S., Anuforo, Ó., Halldórsson, H., Bergmann, A., Halldórsson, S., Palsson, S., Henriksen, H. H., Johansson, P. I. & Rolfsson, Ó. (2018). Metabolic systems analysis of LPS induced endothelial dysfunction applied to sepsis patient stratification. *Sci. Rep.*, *8*(1), 6811. https://doi.org/10.1038/s41598-018-25015-5

Megchelenbrink, W., Huynen, M. & Marchiori, E. (2014). optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. *PLoS One*, *9*(2), e86587. https://doi.org/10.1371/JOURNAL.PONE.0086587

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J. & Barabási, A. L. (2015). Uncovering disease-disease relationships through the incomplete interactome. *Science (80-.).*, *347*(6224), 841. https://doi.org/10.1126/science.1257601

Metallo, C. M., Walther, J. L. & Stephanopoulos, G. (2009). Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. *J. Biotechnol.*, *144*(3), 167. https://doi.org/10.1016/J.JBIOTEC.2009.07.010

Millard, P., Delépine, B., Guionnet, M., Heuillet, M., Bellvert, F. & Létisse, F. (2019). IsoCor: Isotope correction for high-resolution MS labeling experiments. *Bioinformatics*, *35*(21), 4484–4487. https://doi.org/10.1093/bioinformatics/btz209

Millard, P., Letisse, F., Sokol, S. & Portais, J. C. (2012). IsoCor: Correcting MS data in isotope labeling experiments. *Bioinformatics*, *28*(9), 1294–1296. https://doi.org/10.1093/bioinformatics/bts127

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M. & Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. *Nat. Biotechnol.*, *35*(10), 904–908. https://doi.org/10.1038/nbt.3956

Mottaz, A., Yip, Y. L., Ruch, P. & Veuthey, A. L. (2008). Mapping proteins to disease terminologies: From UniProt to MeSH. *BMC Bioinformatics*, *9*(SUPPL. 5). https://doi.org/10.1186/1471-2105-9-S5-S3

Nagana Gowda, G. A. & Djukovic, D. (2014). Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods Mol. Biol.*, *1198*, 3–12. https://doi.org/10.1007/978-1-4939-1258-2_1

Niu, X., Zhang, J., Zhang, L., Hou, Y., Pu, S., Chu, A., Bai, M. & Zhang, Z. (2019). Weighted Gene Co-Expression Network Analysis Identifies Critical Genes in the Development of Heart Failure After Acute Myocardial Infarction. *Front. Genet.*, *0*, 1214. https://doi.org/10.3389/FGENE.2019.01214

Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D. C. & Lewis, N. E. (2017). A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst.*, *4*(3), 318-329.e6. https://doi.org/10.1016/J.CELS.2017.01.010

Orth, J. D., Thiele, I. & Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Ortmayr, K., Dubuis, S. & Zampieri, M. (2019). Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. *Nat. Commun.*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-09695-9

Pacheco, M. P., Bintener, T., Ternes, D., Kulms, D., Haan, S., Letellier, E. & Sauter, T. (2019). Identifying and targeting cancer-specific metabolism with network-based drug target prediction. *EBioMedicine*, *43*, 98–106. https://doi.org/10.1016/j.ebiom.2019.04.046

Palsson, B. O. (2015). Systems biology: Constraint-based reconstruction and analysis. *Syst. Biol. Constraint-Based Reconstr. Anal.*, 1–531. https://doi.org/10.1017/CBO9781139854610

Parini, P., Altucci, L., Balligand, J. L., Baumbach, J., Ferdinandy, P., Filetti, S., Maron, B. A., Petrillo, E., Silverman, E. K., Barabasi, A. L. & Loscalzo, J. (2020). The Network Medicine Imperative and the Need for an International Network Medicine Consortium. *Am. J. Med.*, *133*(9), e451–e454. https://doi.org/10.1016/j.amjmed.2020.03.034

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Stijleman, I. J., Palazzo, J., Marron, J. S., Nobel, A. B., Mardis, E., Nielsen, T. O., Ellis, M. J., Perou, C. M. & Bernard, P. S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.*, *27*(8), 1160–1167. https://doi.org/10.1200/JCO.2008.18.1370

Pasquier, J., Abu-Kaoud, N., Al Thani, H. & Rafii, A. (2015). Epithelial to Mesenchymal Transition in a Clinical Perspective. *J. Oncol.*, *2015*. https://doi.org/10.1155/2015/792182

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., De Clercq, S., Minguijón, E., Balsat, C., Sokolow, Y., Dubois, C., De Cock, F., Scozzaro, S., Sopena, F., Lanas, A., … Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nat. 2018 5567702*, *556*(7702), 463–468. https://doi.org/10.1038/s41586-018-0040-3

Pavlova, N. N. & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.*, *23*(1), 27–47. https://doi.org/10.1016/j.cmet.2015.12.006

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, 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. *Nat. 2000 4066797*, *406*(6797), 747–752. https://doi.org/10.1038/35021093

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.*, *30*(1), 19–34.

Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinforma. 2010 111*, *11*(1), 1–11. https://doi.org/10.1186/1471-2105-11-395

Pramanik, J. & Keasling, J. D. (1998). Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol. Bioeng.*, *60*(2), 230–238.

Price, N. D., Schellenberger, J. & Palsson, B. O. (2004). Uniform Sampling of Steady-State Flux Spaces: Means to Design Experiments and to Interpret Enzymopathies. *Biophys. J.*, *87*(4), 2172–2186. https://doi.org/10.1529/BIOPHYSJ.104.043000

Puisieux, A., Brabletz, T. & Caramel, J. (2014). Oncogenic roles of EMT-inducing transcription factors. *Nat. Cell Biol.*, *16*(6), 488–494. https://doi.org/10.1038/ncb2976

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*.

Redestig, H., Fukushima, A., Stenlund, H., Moritz, T., Arita, M., Saito, K. & Kusano, M. (2009). Compensation for systematic cross-contribution improves normalization of mass spectrometry based metabolomics data. *Anal. Chem.*, *81*(19), 7974–7980. https://doi.org/10.1021/ac901143w

Roessner, U. & Beckles, D. M. (2009). Metabolite measurements. *Plant Metab. Networks*, 39–69. https://doi.org/10.1007/978-0-387-78745-9_3

Rosato, A., Tenori, L., Cascante, M., De Atauri Carulla, P. R., Martins dos Santos, V. A. P. & Saccenti, E. (2018). From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*, *14*(4), 37. https://doi.org/10.1007/s11306-018-1335-y

Rothan, H. A., Stone, S., Natekar, J., Kumari, P., Arora, K. & Kumar, M. (2020). The FDA-approved gold drug auranofin inhibits novel coronavirus (SARS-COV-2) replication and attenuates inflammation in human cells. *Virology*, *547*, 7. https://doi.org/10.1016/J.VIROL.2020.05.002

Rusilowicz, M., Dickinson, M., Charlton, A., O'Keefe, S. & Wilson, J. (2016). A batch correction method for liquid chromatography–mass spectrometry data that does not depend on quality control samples. *Metabolomics*, *12*(3), 1–11. https://doi.org/10.1007/S11306-016-0972-2

Schellenberger, J. & Palsson, B. (2009). Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.*, *284*(9), 5457–5461. https://doi.org/10.1074/jbc.R800048200

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. (2016). Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.*, *27*(12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Schultz, A. & Qutub, A. A. (2016). Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLoS Comput. Biol.*, *12*(3). https://doi.org/10.1371/journal.pcbi.1004808

Sciacovelli, M., Gonçalves, E., Johnson, T. I., Zecchini, V. R., Costa, A. S. H. da, Gaude, E., Drubbel, A. V., Theobald, S. J., Abbo, S., Tran, M., Rajeeve, V., Cardaci, S., Foster, S., Yun, H., Cutillas, P., Warren, A., Gnanapragasam, V., Gottlieb, E., Franze, K., … Frezza, C. (2016). Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature*, *537*(7621), 544. https://doi.org/10.1038/NATURE19353

Senan, O., Aguilar-Mogas, A., Navarro, M., Capellades, J., Noon, L., Burks, D., Yanes, O., Guimera, R. & Sales-Pardo, M. (2019). CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics*, *35*(20), 4089–4097. https://doi.org/10.1093/bioinformatics/btz207

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. https://doi.org/10.1016/j.cell.2014.07.032

Shibue, T. & Weinberg, R. A. (2017). EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.*, *14*(10), 611–629. https://doi.org/10.1038/nrclinonc.2017.44

Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, *6*(10), 813–823. https://doi.org/10.1038/nrc1951

Shubham, K., Vinay, L. & Vinod, P. K. (2017). Systems-level organization of non-alcoholic fatty liver disease progression network. *Mol. Biosyst.*, *13*(9), 1898–1911. https://doi.org/10.1039/c7mb00013h

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. https://doi.org/10.1371/journal.pone.0023833

Silverman, E. K., Schmidt, H. H. H. W., Anastasiadou, E., Altucci, L., Angelini, M., Badimon, L., Balligand, J. L., Benincasa, G., Capasso, G., Conte, F., Di Costanzo, A., Farina, L., Fiscon, G., Gatto, L., Gentili, M., Loscalzo, J., Marchese, C., Napoli, C., Paci, P., … Baumbach, J. (2020). Molecular networks in Network Medicine: Development and applications. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, *12*(6), e1489. https://doi.org/10.1002/wsbm.1489

Sinclair, C., Bommakanti, G., Gardinassi, L., Loebbermann, J., Johnson, M. J., Hakimpour, P., Hagan, T., Benitez, L., Todor, A., Machiah, D., Oriss, T., Ray, A., Bosinger, S., Ravindran, R., Li, S. & Pulendran, B. (2017). mTOR regulates metabolic adaptation of APCs in the lung and controls the outcome of allergic inflammation. *Science (80-.).*, *357*(6355), 1014–1021. https://doi.org/10.1126/SCIENCE.AAJ2155

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. *Anal. Chem.*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Stokvis, E., Rosing, H. & Beijnen, J. H. (2005). Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: Necessity or not? *Rapid Commun. Mass Spectrom.*, *19*(3), 401–407. https://doi.org/10.1002/rcm.1790

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sysi-Aho, M., Katajamaa, M., Yetukuri, L. & Orešič, M. (2007). Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics*, *8*(1), 1–17. https://doi.org/10.1186/1471-2105-8-93

Széliová, D., Ruckerbauer, D. E., Galleguillos, S. N., Petersen, L. B., Natter, K., Hanscho, M., Troyer, C., Causon, T., Schoeny, H., Christensen, H. B., Lee, D. Y., Lewis, N. E., Koellensperger, G., Hann, S., Nielsen, L. K., Borth, N. & Zanghellini, J. (2020). What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metab. Eng.*, *61*, 288–300. https://doi.org/10.1016/j.ymben.2020.06.002

Takemoto, K. (2014). Metabolic networks are almost nonfractal: A comprehensive evaluation. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, *90*(2). https://doi.org/10.1103/PhysRevE.90.022802

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 α Is a Central Signaling Node and Therapeutic Target for Breast Cancer Stem Cells. *Cancer Cell*, *24*(3), 347–364. https://doi.org/10.1016/j.ccr.2013.08.005

Tanaka, R. (2005). Scale-rich metabolic networks. *Phys. Rev. Lett.*, *94*(16). https://doi.org/10.1103/PhysRevLett.94.168101

Tang, W., Zhou, M., Dorsey, T. H., Prieto, D. A., Wang, X. W., Ruppin, E., Veenstra, T. D. & Ambs, S. (2018). Integrated proteotranscriptomics of breast cancer reveals globally increased protein-mRNA concordance associated with subtypes and survival. *Genome Med.*, *10*(1), 94. https://doi.org/10.1186/s13073-018-0602-x

Tarragó-Celada, J., Foguet, C., Tarrado-Castellarnau, M., Marin, S., Hernández-Alias, X., Perarnau, J., Morrish, F., Hockenbery, D., Gomis, R. R., Ruppin, E., Yuneva, M., Atauri, P. de & Cascante, M. (2021). Cysteine and Folate Metabolism Are Targetable Vulnerabilities of Metastatic Colorectal Cancer. *Cancers 2021, Vol. 13, Page 425*, *13*(3), 425. https://doi.org/10.3390/CANCERS13030425

The Icelandic Cancer Society. (2021). *Breast*. https://www.krabb.is/krabbameinsskra/en/cancer-sites/cancers-a-z/breast/

The World Health Organization. (2021). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer

Thiele, I. & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, *5*(1), 93. https://doi.org/10.1038/NPROT.2009.203

Thiele, I., Price, N. D., Vo, T. D. & Palsson, B. (2005). Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J. Biol. Chem.*, *280*(12), 11683–11695. https://doi.org/10.1074/jbc.M409072200

Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., Thorleifsson, S. G., Agren, R., Bölling, C., Bordel, S., Chavali, A. K., Dobson, P., Dunn, W. B., Endler, L., Hala, D., … Palsson, B. O. (2013). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, *31*(5), 419–425. https://doi.org/10.1038/nbt.2488

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. *Nat. Methods 2015 126*, *12*(6), 523–526. https://doi.org/10.1038/nmeth.3393

Uppal, K., Walker, D. I. & Jones, D. P. (2017). xMSannotator: An R package for network-based annotation of high-resolution metabolomics data. *Anal. Chem.*, *89*(2), 1063–1067. https://doi.org/10.1021/acs.analchem.6b01214

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 142. https://doi.org/10.1186/1471-2164-7-142

Van Winden, W. A., Wittmann, C., Heinzle, E. & Heijnen, J. J. (2002). Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol. Bioeng.*, *80*(4), 477–479. https://doi.org/10.1002/bit.10393

Villadsen, R., Fridriksdottir, A. J., Rønnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M. A., Bissell, M. J. & Petersen, O. W. (2007). Evidence for a stem cell hierarchy in the adult human breast. *J. Cell Biol.*, *177*(1), 87–101. https://doi.org/10.1083/jcb.200611114

Vlassis, N., Pacheco, M. P. & Sauter, T. (2014). Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLoS Comput. Biol.*, *10*(1), e1003424. https://doi.org/10.1371/journal.pcbi.1003424

Wang, L., Hou, E., Wang, L., Wang, Y., Yang, L., Zheng, X., Xie, G., Sun, Q., Liang, M. & Tian, Z. (2015). Reconstruction and analysis of correlation networks based on GC-MS metabolomics data for young hypertensive men. *Anal. Chim. Acta*, *854*, 95–105. https://doi.org/10.1016/j.aca.2014.11.009

Wang, S., Cyronak, M. & Yang, E. (2007). Does a stable isotopically labeled internal standard always correct analyte response?: A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. *J. Pharm. Biomed. Anal.*, *43*(2), 701–707. https://doi.org/10.1016/J.JPBA.2006.08.010

Wang, Y., Eddy, J. A. & Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol. 2012 61*, *6*(1), 1–16. https://doi.org/10.1186/1752-0509-6-153

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., … Scalbert, A. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.*, *46*(D1), D608–D617. https://doi.org/10.1093/nar/gkx1089

Wong, C. C., Qian, Y. & Yu, J. (2017). Interplay between epigenetics and metabolism in oncogenesis: Mechanisms and therapeutic approaches. *Oncogene*, *36*(24), 3359–3374. https://doi.org/10.1038/onc.2016.485

Worley, B. & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Curr. Metabolomics*, *1*(1), 92–107. https://doi.org/10.2174/2213235x11301010092

Worley, B. & Powers, R. (2016). PCA as a practical indicator of OPLS-DA model reliability. *Curr. Metabolomics*, *4*(2), 97. https://doi.org/10.2174/2213235X04666160613122429

Xia, J. & Wishart, D. S. (2010). MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.*, *38*(SUPPL. 2), W71–W77. https://doi.org/10.1093/nar/gkq329

Xia, J., Wishart, D. S. & Valencia, A. (2011). MetPA: A web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*, *27*(13), 2342–2344. https://doi.org/10.1093/bioinformatics/btq418

Yaneske, E. & Angione, C. (2017). A Data- and Model-Driven Analysis Reveals the Multi-omic Landscape of Ageing. *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, *10208 LNCS*, 145–154. https://doi.org/10.1007/978-3-319-56148-6_12

Yang, L., Li, Y., Wei, Z. & Chang, X. (2018). Coexpression network analysis identifies transcriptional modules associated with genomic alterations in neuroblastoma. *Biochim. Biophys. Acta - Mol. Basis Dis.*, *1864*(6), 2341–2348. https://doi.org/10.1016/J.BBADIS.2017.12.020

Zampieri, G., Coggins, M., Valle, G., Angione, C., Zampieri, G., Coggins, M., Valle, G. & Angione, C. (2017). *A poly-omics machine-learning method to predict metabolite production in CHO cells*. 4993. https://doi.org/10.3390/IECM-2-04993

Zampieri, G., Vijayakumar, S., Yaneske, E. & Angione, C. (2019). Machine and deep learning meet genome-scale metabolic modeling. *PLOS Comput. Biol.*, *15*(7), e1007084. https://doi.org/10.1371/JOURNAL.PCBI.1007084

Zhang, J., Tian, X.-J. & Xing, J. (2016). Signal Transduction Pathways of EMT Induced by TGF-β, SHH, and WNT and Their Crosstalks. *J. Clin. Med.*, *5*(4), 41. https://doi.org/10.3390/jcm5040041

Zhang, K., Mack, P. & Wong, K. P. (1998). Glutathione-related mechanisms in cellular resistance to anticancer drugs (Review). *Int. J. Oncol.*, *12*(4), 871–882. https://doi.org/10.3892/ijo.12.4.871

Zhang, Z., Chen, L., Xu, P., Xing, L., Hong, Y. & Chen, P. (2020). Gene correlation network analysis to identify regulatory factors in sepsis. *J. Transl. Med. 2020 181*, *18*(1), 1–15. https://doi.org/10.1186/S12967-020-02561-Z

Zhao, Y., Tian, B., Sadygov, R. G., Zhang, Y. & Brasier, A. R. (2016). Integrative proteomic analysis reveals reprograming tumor necrosis factor signaling in epithelial mesenchymal transition. *J. Proteomics*, *148*, 126–138. https://doi.org/10.1016/J.JPROT.2016.07.014

Zhou, Y., Hou, Y., Shen, J., Mehra, R., Kallianpur, A., Culver, D. A., Gack, M. U., Farha, S., Zein, J., Comhair, S., Fiocchi, C., Stappenbeck, T., Chan, T., Eng, C., Jung, J. U., Jehi, L., Erzurum, S. & Cheng, F. (2020). A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLoS Biol.*, *18*(11), e3000970. https://doi.org/10.1371/journal.pbio.3000970

Zhu, J., Berisa, M., Schwörer, 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.e5. https://doi.org/10.1016/J.CMET.2019.09.009

Zielinski, D. C., Jamshidi, N., Corbett, A. J., Bordbar, A., Thomas, A. & Palsson, B. O. (2017). Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Sci. Rep.*, *7*(1), 1–14. https://doi.org/10.1038/srep41241

Original publications

Paper I

EMT-Derived Alterations in Glutamine Metabolism Sensitize Mesenchymal Breast Cells to mTOR Inhibition

Sigurdur Trausti Karvelsson[1], Arnar Sigurdsson[2], Kotryna Seip[3], Maria Tunset Grinde[4], Qiong Wang[1], Freyr Johannsson[1], Gunhild Mari Mælandsmo[3], Siver Andreas Moestue[5,6], Ottar Rolfsson[1], and Skarphedinn Halldorsson[1,7]

ABSTRACT

Epithelial-to-mesenchymal transition (EMT) is a fundamental developmental process with strong implications in cancer progression. Understanding the metabolic alterations associated with EMT may open new avenues of treatment and prevention. Here we used ^{13}C carbon analogs of glucose and glutamine to examine differences in their utilization within central carbon and lipid metabolism following EMT in breast epithelial cell lines. We found that there are inherent differences in metabolic profiles before and after EMT. We observed EMT-dependent re-routing of the TCA-cycle, characterized by increased mitochondrial IDH2-mediated reductive carboxylation of glutamine to lipid biosynthesis with a concomitant lowering of glycolytic rates and glutamine-dependent glutathione (GSH) generation. Using weighted correlation network analysis, we identified cancer drugs whose efficacy against the NCI-60 Human Tumor Cell Line panel is significantly associated with GSH abundance and confirmed these *in vitro*. We report that EMT-linked alterations in GSH synthesis modulate the sensitivity of breast epithelial cells to mTOR inhibitors.

Implications: EMT in breast cells causes an increased demand for glutamine for fatty acid biosynthesis, altering its contribution to glutathione biosynthesis, which sensitizes the cells to mTOR inhibitors.

Introduction

Epithelial-to-mesenchymal transition (EMT) is a fundamental developmental process where tightly bound epithelial cells differentiate into migratory mesenchymal cells that can relocate into adjacent or distant tissues. This process is vital for tissue restructuring during embryonic development and is also necessary for proper wound healing in adult tissue. EMT has strong implications in cancer progression and metastasis where primary tumor cells of epithelial origin can take on a motile phenotype with the ability to migrate through the body and establish secondary tumors at distant locations (1).

Metabolic reprogramming is recognized as one of the 10 cancer hallmarks as proposed by Hanahan and Weinberg (2). In contrast to rapidly dividing cancer cells, a mesenchymal phenotype faces a different set of metabolic requirements whose relation to malignant transformation has been intensely studied and associated with enhanced glycolysis, increased glutaminolysis, nucleotide metabolism, and abnormal choline metabolism (3–5). Quantitative understanding of the metabolic requirements of mesenchymal cells is however lacking, particularly the changes to the turnover and quantity of metabolites involved in xenobiotic clearance, that is, the drug response of cells. Cancer cells that undergo EMT have increased resistance to various drugs (6–8), which indicates that the xenobiotic clearance of the cells is altered. There are three phases involved in the metabolism of xenobiotics: (i) modification, (ii) conjugation, and (iii) excretion. Conjugation involves the binding of particular metabolites (e.g., glutathione (GSH), UDP-glucuronate, PAPS, S-adenosylmethionine) to a xenobiotic compound (9), which leads to the assumption that the availability of these metabolites within cells influences the activity of the drugs. Therefore, accurate metabolic measurements of EMT may contribute to better understanding of the drug resistance of cancer cells and lead to novel therapeutic approaches aimed at eliminating metastatic cancer cells.

We have previously used both ultra-performance liquid chromatography coupled mass spectrometry (UPLC-MS) and NMR to study EMT and cancer metabolism (10–12). Integrated analyses of these metabolomics data with transcriptomic and proteomic data within genome-scale metabolic models predicted metabolic differences that occur following EMT in breast epithelium (12). These included alterations to glycolysis, the pentose phosphate pathway, TCA cycle, and fatty acid synthesis. Although these models provided useful insights into metabolic alterations associated with EMT, they lacked accuracy in predicting internal fluxes in a quantitative manner in the compartmentalized central carbon metabolism.

To better understand the metabolic changes that accompany EMT, we characterized the internal flow of metabolites in D492 breast epithelial cells and their mesenchymal variant, D492M, to determine metabolic changes within central carbon metabolism following EMT in breast epithelial cells. We performed stable isotope tracing of ^{13}C labeled glucose and two separate ^{13}C labeled glutamine analogs. UPLC-MS and NMR were used to measure label incorporation into metabolites associated with central carbon metabolism and lipid biosynthesis. We subsequently performed shRNA lentiviral silencing of key genes to further elucidate their role in EMT metabolic reprogramming. Finally, using an integrated network analysis of the

[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland. [2]Department of Chemistry, Technische Universität Berlin, Berlin, Germany. [3]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. [4]Department of Circulation and Medical Imaging, NTNU, Trondheim, Norway. [5]Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway. [6]Department of Pharmacy, Nord University, Namsos, Norway. [7]Institute for Surgical Research, Vilhelm Magnus Laboratory, Oslo University Hospital, Oslo, Norway.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).

Corresponding Author: Óttar Rolfsson, Center for Systems Biology, University of Iceland, Sturlugata 8, Reykjavik 101, Iceland. Phone: 354-845-0075; E-mail: ottarr@hi.is

Mol Cancer Res 2021;19:1546–58

doi: 10.1158/1541-7786.MCR-20-0962

NCI-60 Human Tumor Cell Line panel and an untargeted metabolomic analysis, we investigate how the EMT-dependent re-routing of central carbon metabolism affects drug responsiveness in D492 and D492M cells.

Materials and Methods

Cell culture

D492 and D492M cells were kindly provided by the Stem Cell Research Unit, University of Iceland, and were cultured in DMEM/F12-based medium H14 at 37°C in 5% CO_2 as described previously (13). All experiments were performed within four passages from thawing, within the range of 30 to 40 passages in total. For the labeling experiments, the cells were fed with medium containing 100% ^{13}C-labeled glutamine at the one or five position (Cambridge Isotope Laboratories, Inc.) or ^{13}C-labeled glucose at the one and two positions (Cambridge Isotope Laboratories, Inc.). Cells were screened for *Mycoplasma* infections every month using PCR-based tests at the Biomedical Center, University of Iceland.

Lentiviral shRNA production and transduction

HEK293T cells were transfected using TurboFect transfection reagent (Thermo Fisher Scientific) at 80% confluency in T25 cell culture vessels. The cells were then incubated at 37°C and 5% CO_2. Viral supernatant was collected at two timepoints, the first being after 48 hours in culture, and the second 72 hours after changing medium at the first timepoint. The viral supernatant was filtered through a 0.45 μm filter using a syringe and stored at −20°C until usage. The lentiviral vectors were acquired from GeneCopoeia. They contained an shRNA construct for the selective targeting of IDH2. The construct was based on a psi-LVRH1MH vector with an mCherry fluorescent reporter, resistance against hygromycin B, and the identical hairpin sequence TCAAGAG. The target sequence was *IDH2* 5′GTACAAGGCCACAGACTTTGT-3′. The D492 and D492M cell lines were transduced using 1 mL of filtered viral supernatant at 70% confluency and incubated at 37°C and 5% CO_2 for 24 hours, at which timepoint the medium was changed to fresh H14 medium. After further 48 hours, the cells were grown in medium containing hygromycin B (200 μg/mL) for 3 weeks to selectively grow cells containing the shRNA construct.

Real-time PCR

Whole-cell RNA was extracted using Tri-Reagent (Thermo Fisher Scientific, AM9738). Reverse transcription was performed using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, 4368814). The expression of the genes *IDH1* and *IDH2* was quantified, where *ACTB* (Beta-actin) and was used as an endogenous reference gene. The primers for *IDH1*, *IDH2*, and *ACTB* were designed using the Primer3 software in the Benchling website (https://benchling.com). Primers were designed to span exon junctions and have a melting temperature above 55°C. The expression of *IDH2* and *IDH1* was assessed using real-time PCR (qPCR). Real-time quantitative PCR reactions were carried out using Luna Universal qPCR Master Mix (New England Biolabs) according to manufacturer's instructions on a Bio-Rad CFX384 Touch Real Time System (Bio-Rad). Gene expression levels were determined using CFX Manager Software (Bio-Rad) and differences in relative expression were estimated with the $2^{\Delta\Delta Ct}$ method. The primer sequences used for quantifying the gene expression were: *IDH1*-forward 5′-CGACATGGTGGCCCAAGCTATG-3′, *IDH1*-reverse 5′-TCATGCCGAGAGAGCCATACCC-3′, *IDH2*-forward 5′-ATGAGGCCCGTGTGGAAGAGTT-3′, *IDH2*-reverse 5′-CAGATGATGGGCTCCCGGAAGA-3′, ACTB-forward 5′-CTTCCTGGGTGAGTGGAGACTG-3′, and ACTB-reverse 5′-GAGGGAAATGAGGGCAGGACTT-3′.

Proliferation assay

Cells were seeded in quadruplicates in 48-well plates (10,000 cells/well). They were grown in a large chamber incubation system (PeCon GmbH) at 37°C in 5% CO_2 and imaged for 12 to 72 hours using Leica DMI6000B. Images of cells were opened with Fiji (14), where the cells were counted with the help of an in-house script.

Detection of intracellular NADP$^+$ and NADPH

NADP+ and NADPH were measured using NADP/NADPH-Glo Assay (G9081; Promega). Cells were seeded in triplicates in opaque 96-well plates (10,000 cells/well) and incubated at 37°C in 5% CO_2. After 24 hours, the medium was removed, cells were washed with cold PBS, and then supplemented with 50 μL PBS and 50 μL 1% DTAB in 0.2N NaOH solution to induce cell lysis. Next steps were according to manufacturer's protocol. The luminescence was measured 50 minutes after addition of the NADP-NADPH-Glo Detection reagent in SpectraMaxM3 Microplate Reader (Molecular Devices).

Nuclear magnetic resonance (NMR)

For NMR analysis, D492 and D492M cells were cultured in T225 flasks in supplemented DMEM/F12 until they reached approximately 70% confluency. Cells were then fed with either 1,2–^{13}C glucose or 1–^{13}C glutamine for 6 hours. Parallels without ^{13}C tracers were also cultured. Culture medium was collected after incubation. Methanol extracts from glucose- and glutamine-labeled cells were prepared as described previously (15). The cell extracts were freeze dried prior to NMR analysis. For NMR, freeze-dried cell extracts were dissolved in 600 μL D_2O in PBS whereas culture medium (500 μL) was diluted with in D_2O-based PBS (100 μL). NMR analysis was performed using a 600 MHz Bruker Avance III NMR spectrometer (Bruker Biospin GmbH), equipped with a 5 mm QCI Cryoprobe with integrated, cooled preamplifiers for ^{1}H, ^{2}H, and ^{13}C. Proton spectra were acquired at 300 K using 1D NOESY (Bruker: noesygppr1d) with presaturation and spoiler gradients as described previously (16). The spectra were collected with 32 scans and 4 dummy scans. The acquisition time was 2.73 seconds and relaxation delay 4 seconds, measuring the FID via collection of 64 K complex data points. The ^{1}H spectra were Fourier transformed with a 0.3 Hz exponential line broadening and the chemical shift was calibrated to alanine at 1.48 ppm. ^{1}H spectra from D492 ($n = 5$) and D492M ($n = 6$) cells were transferred to MATLAB R2017a for multivariate data analysis. The spectra were baseline corrected using asymmetric least squares method (17) and peak aligned using icoshift (18). The water peak and areas in the spectra with contamination and noise only were removed. All spectra were mean normalized and mean centered. Principal component analysis (PCA) was performed using PLS toolbox v8.2.1 (Eigenvector Research). Proton decoupled ^{13}C spectra (Bruker: zgpg30) were acquired using a power gated decoupling sequence with a 30° pulse angle as described in Bettum and colleagues (19). The spectra were collected with either 4 K (for 1,2–^{13}C-glucose) or 16 K (for 1–^{13}C-glutamine) scans and 16 dummy scans. The acquisition time was 1.65 seconds, relaxation delay 0.5 seconds, measuring the FID via collection of 96 K complex data points over a sweep width of 197.175 ppm. The ^{13}C spectra were Fourier transformed with a 3.0 Hz exponential line broadening and the chemical shift was calibrated to the 3–^{13}C-alanine peak at 19.0 ppm or 1–^{13}C-glutamine peak at 176.4 ppm. ^{13}C-labeled metabolites downstream from the tracers were identified by

comparing ^{13}C spectra with natural abundance spectra acquired under the same conditions. Levels of selected metabolites in the extracts were semiquantitatively assessed by integration of resonance signals using TopSpin 4.0.8 (Bruker Biospin GmbH) after correcting for natural abundance levels. The ^{13}C spectra were normalized to the total AUC in the ^{1}H spectra acquired from the same sample.

Metabolomics

Sample extraction

Polar and nonpolar metabolites were extracted from cell cultures by methanol/chloroform/water extraction. Cells were harvested in ice-cold methanol, vortexed vigorously, and let stand on ice for 10 minutes. Equal amounts of water and chloroform were added to a final composition of 1:1:1 (CH_3OH:H_2O:$CHCl_3$), vortexed, and left to stand overnight at 4°C. The organic phase (lipids) was collected into a glass vial and dried in a stream of N_2 and stored under N_2 at −80°C until analysis. The aqueous phase (polar metabolites) was stored at −80°C and evaporated in a miVac concentrator (SP Scientific) before analysis.

UPLC-MS

Before UPLC-MS analysis, the organic phase was reconstituted in MTBE before a methanol solution containing 1M NaOH was added (10:1 v/v, respectively). This was incubated for 3.5 hours at 37°C, when 1 μL formic acid was added (to neutralize the solution), the samples were dried in a stream of N_2 and then resuspended in isopropanol:ACN:H_2O (6:9:1, v/v/v). The aqueous phase (metabolites) were reconstituted in isopropanol:ACN:H_2O (2:1:1, v/v/v). UPLC (Acquity) was coupled with a quadrupole-time of flight mass spectrometer (Synapt G2; Waters). For the lipid samples, chromatographic separation was achieved as described previously (20). For the metabolomic samples, chromatographic separation was achieved by hydrophilic interaction liquid chromatography using an Acquity amide column, 1.7 μm (2.1 × 150 mm; Waters). All samples were analyzed in positive ionization and negative ionization mode using acidic and basic chromatographic conditions. In positive mode and in negative acidic conditions, mobile phase A was 100% ACN and B was 100% H2O both containing 0.1% formic acid. The following elution gradient was used: 0 minutes 99% A; 7 minutes 30% A; 7.1 minutes 99% A; 10 minutes 99% A. In negative mode basic conditions, mobile phase A contained ACN:sodium bicarbonate 10 mmol/L (95:5) and mobile phase B contained ACN: sodium bicarbonate 10 mmol/L (5:95). The following elution gradient was used: 0 minute 99% A; 6 minutes 30% A; 6.5 minutes 99% A; 10 minutes 99% A. In all conditions, the flow rate was 0.4 mL/min, the column temperature was 45°C, and the injection volume was 3.5 μL. The mass spectrometer was operated using a capillary voltage of 1.5 kV, the sampling cone and the extraction cone were of 30 and 5 V. The cone and the desolvation gas flow were 50 and 800 L/h, whereas the source and desolvation gas temperature were 120°C and 500°C. MS spectra were acquired in centroid mode from m/z 50 to 1,000 using scan time of 0.3 seconds. Leucine enkephalin (2 ng/μL) was used as lock mass (m/z 556.2771 and 554.2615 in positive and negative experiments, respectively). A typical analytical block consisted of: (i) pooled QC samples to equilibrate the system, (ii) calibrators, (iii) samples and spiked pooled QC samples, and (iv) calibrators.

Data analysis

TargetLynx (v4.1; Waters) was used to integrate chromatograms of all isotopologues of the metabolites of interest. Ion chromatograms were extracted using a window of 0.02 mDa, which was centered on the expected m/z for each targeted isotopologue. The output was a mass distribution vector (MDV) describing the relative amount of each detected isotopologue of the metabolite. Ion chromatograms of isotopologues of interest extracted and corrected for abundance of naturally occurring isotopes using the IsoCor software (21). When calculating the total contribution (TC) of carbon sources to metabolites, we used the following equation (22):

$$TC = \frac{\sum_{i=0}^{n} im_i}{n} \quad (1)$$

where n is the number of C atoms in the metabolite, i represents the isotopologues, and m is the relative fraction of the isotopologues.

To evaluate the percentage of glucose that enters the oxidative part of the pentose phoshate pathway, and re-enters glycolysis, we utilized a formula from Lee and colleagues (23):

$$PPP_{Cycle} = \frac{m1/m2}{3 + m1/m2} \quad (2)$$

In Equation (2), m1 and m2 are the fractional abundances of M+1 and M+2 lactate isotopologues, respectively (e.g., from Supplementary Fig. S2).

RNA sequencing

Quantified transcript pseudocounts from kallisto (24) were obtained for D492 and D492M in triplicates from Briem and colleagues (from the authors; ref. 25). These data were imported into R and simultaneously log2-transformed and variance-stabilized using DESeq2′s rlog function (26).

Proteomic analysis

A proteomic dataset for the D492 and D492M cells was obtained from the ProteomeXchange Consortium via the PRIDE (27) partner repository with the dataset identifier PXD024164. The raw data were processed using MaxQuant (28) for both the protein identification and quantification.

Western blot analysis

D492 and D492M cells were grown to 80% to 90% confluent as described above followed by lysis in RIPA buffer. The lysates were subjected to five freeze–thaw cycles, centrifuged at 14,000 RCF for 20 minutes at 4°C. Protein concentration was quantified using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).

10 to 20 μg of protein were loaded onto precast 4% to 12% NuPAGE Bis-Tris gels (Thermo Fisher Scientific) and transferred to a 0.45 μm nitrocellulose membrane (Thermo Fisher Scientific). The membrane was blocked in 5% BSA (Thermo Fisher Scientific) for 60 minutes, followed by overnight primary antibody incubation at 4°C. The primary antibody was anti-IDH2 monoclonal rabbit (12652; Cell Signaling Technologies) in 1:1,000 dilution. IDH2 levels were normalized against β-actin (MA5–15739; Thermo Fisher Scientific) in 1:10,000 dilution.

Bands were detected by secondary antibody incubation for 2 hours at room temperature using anti-rabbit IgG (H+L) DyLight 800 4× PEG Conjugate and anti-mouse IgG (H+L) DyLight 680 Conjugate (Cell Signaling Technologies) in 1:15,000 dilution. Imaging was performed using Odyssey CLx (LI-COR Biosciences) to scan the films at 169 μm resolution. The results were analyzed in the Image Lab software (Bio-Rad).

Measurement of metabolite exchange rates

Glucose and lactate measurements were performed in an ABL 90 blood gas analyzer (Radiometer). Glutamine uptake was measured using L-Glutamine/Ammonia Assay Kit (K-GLNAM; Megazyme). The following formula was used to quantify the exchange rates of metabolites in the cells:

$$v_k = \frac{V\left([M_k]_f - [M_k]_i\right)}{A} \quad (3)$$

where v_k is the exchange rate for metabolite k, $[M_k]_{48}$ and $[M_k]_0$ are the concentrations of metabolite k in the culture media after 48 and 0 hours, respectively, and A is the area under the growth curve.

Weighted correlation network analysis

Drug sensitivity data were gathered from the NCI-60 Human Tumor Cell Lines (29) using the rcellminer R-package (30). We focused specifically on FDA-approved drugs, so after filtering the data (using WGCNA's goodSamplesGenes function; ref. 31), the final size of the drug sensitivity matrix was 214 drugs × 59 cell lines (one cell line, MDA-N, was removed due to missing data). To construct a network of drug-sensitivities, a weighted correlation network analysis was employed using the WGCNA R-package (31). A soft threshold value of 9 was used to obtain a scale-free network topology ($R^2 = 0.85$). Highly correlated modules that had an average distance <0.65 were merged. To associate the modules to a specific mechanism of action (MOA), the MOA terms for each drug were obtained (from rcellminer) and tested for overrepresentation in the modules. MOA with Bonferroni-adjusted P-value <0.05 were labeled as overrepresented and used for the functional annotation of the drug modules.

The eigenvalues (first principal components) for each module were identified and used to test association of drug-modules with metabolite levels. The correlation of the drug modules to metabolite levels of the cells within the NCI-60 (from ref. 32) was calculated. The R-package igraph (33) was used to visualize the drug correlation network using the Fruchterman-Reingold (34) force-directed layout. An R-script for the network analysis is in Supplementary File S1.

Drug treatment assays

Cells were seeded (3,000 cells/well) in white Costar 96-well plates (Corning) and maintained at 37°C in 5% CO_2. After 24 hours, drugs were added with or without buthionine sulphoximine (BSO; B2515; Sigma Aldrich) where DMSO was used as a control. Then, 72 hours later cell viability was evaluated using the CellTiter-Glo (CTG) assay (Promega), by adding CTG assay mix directly into the wells in a 1:1 ratio. After 10 minutes, luminescence was measured by Victor X3 Multiplate reader (Perkin Elmer). The drugs tested were mTOR inhibitor everolimus (Sigma-Aldrich) and taxane drug paclitaxel (Fresenius Kabi).

Detection of intracellular GSH abundance

Cells were seeded (2,000 cells/well) in white 384-well plates (Greiner Bio-One) and maintained at 37°C in 5% CO_2. After 24 hours, BSO (treated) or medium (nontreated) was added. After additional 24 (and 48 for nontreated) hours, GSH measurement was performed by using GSH-Glo Glutathione Assay Kit (V6911; Promega) in accordance with the manufacturer's protocol.

Statistical analysis

Student t test was employed for comparison of two treatments. Benjamini–Hochberg adjustment for multiple comparisons was performed when appropriate. ANOVA was used to compare data from three or more treatments or the simultaneous evaluation of the effect of two grouping variables. The asterisks in each figure represent the P values ($^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; ns, not significant). Data were assumed to be normally distributed. Statistical analysis and image generation was carried out in the R environment (35) using the *ggplot2* (36) and *ggpubr* (37) packages. In our graphs, all data points are plotted and summarized using mean and SE.

Results

Glycolysis rates determine the pentose phosphate shunt in the D492 EMT cell model

A clear difference in the overall metabolic profiles of D492 and D492M cells was confirmed by PCA of their 1H NMR spectra. The score and loading plots from PCA (Supplementary Figs. S1A and S1B) indicated that D492 cells had more intracellular isoleucine, leucine, valine, alanine, arginine, GSH, myo-inositol, asparagine, proline, AMP, ADP, ATP, tyrosine, phenylalanine, and NAD+, and less glutamine, glutamate, phosphocholine, glycine, threonine, glucose, fumarate, NADP, and NADH. The rate of glucose uptake and lactate secretion was also higher in the epithelial phenotype of D492 cells compared with the mesenchymal phenotype (**Fig. 1B**), indicative of enhanced glycolysis in D492. To determine pentose phosphate pathway (PPP) split ratios in the cell lines, we used 1,2–^{13}C glucose (**Fig. 1A**), as described previously (38). Label contribution of 1,2–^{13}C-glucose to lactate after 6 hours was higher in D492 epithelial cells than in D492M mesenchymal cells (**Fig. 1C**), confirming enhanced glycolysis. To determine the differences in PPP activity, we calculated the percentage of glucose diverted into the PPP using measured lactate isotopologue abundances (Supplementary Fig. S2) and Equation (2). Roughly 2% of glucose was found to enter the PPP cycle in both cell lines. Therefore, because of the overall higher glucose uptake, the flux into oxidative PPP is higher in D492. As a result, reduction of NADP to NADPH via the oxidative phase of the PPP in D492M cells is dampened.

To examine the contribution of glucose to the TCA cycle, we measured the contribution of the labeled 1,2–^{13}C-glucose to citrate. Citrate is either oxidized for energy production in the TCA cycle or used as a precursor for lipid biosynthesis (39) through ATP-citrate lyase (ACLY) in the cytosol (**Fig. 1A**). No difference in glucose-dependent citrate generation (**Fig. 1D**) was observed, which was reflected by the contribution of glucose to palmitate (**Fig. 1E**). Furthermore, the contribution of glucose to other TCA cycle components and downstream metabolites (malate, aspartate, and glutamate) after 6 hours in culture was minimal (Supplementary Fig. S2). Thus, the only difference in glucose utilization within central carbon metabolism between D492 and D492M was the increased glycolytic activity of the former.

Glutamine fuels citrate and lipogenic acetyl-CoA production via reductive carboxylation following EMT in D492 cells

Glutamine is a major contributor of carbons into the TCA cycle through anaplerosis (i.e., glutaminolysis), particularly in cancer cells (40–42). Glutamine is the second-highest consumed carbon source in D492 and D492M (after glucose; ref. 12) with an average uptake rate of around 60 fmoles/cell/hour in both cell lines (**Fig. 2B**). We found that both glucose and glutamine are essential for the proliferation of D492 and D492M (Supplementary Fig. S3). Glutamine can replenish the TCA cycle via glutamate and α-ketoglutarate, which can be metabolized within the TCA cycle both oxidatively and

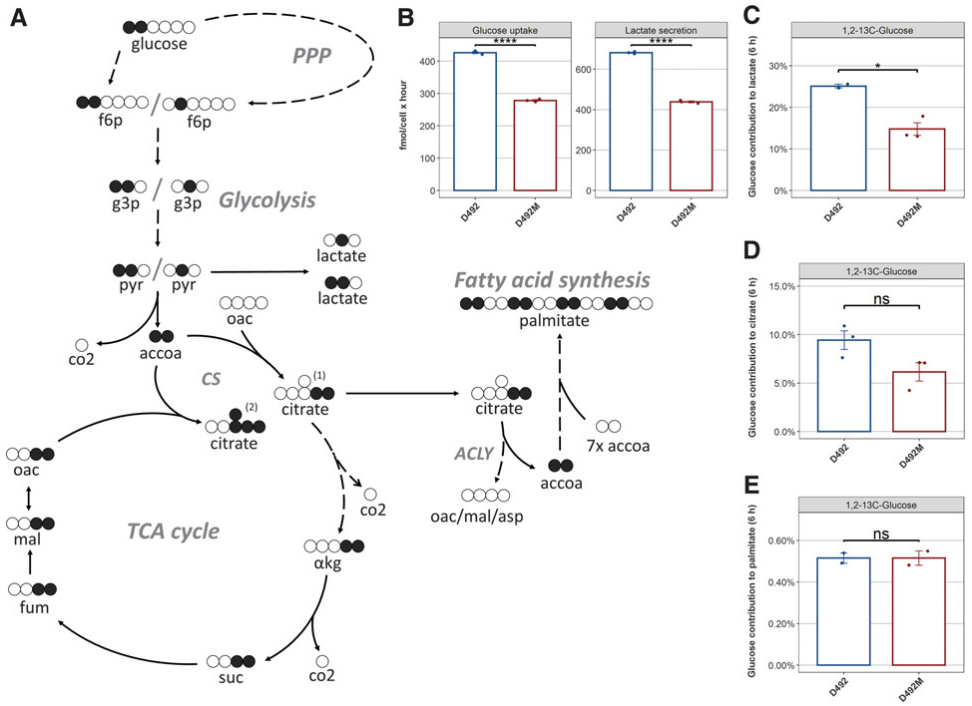


Figure 1.

Glucose metabolism of D492 and D492M. **A,** A schematic overview of label distribution from 1,2-^{13}C glucose into central carbon metabolism. An atom transition map of glucose metabolism showing a part of the metabolic fates of 1,2-^{13}C glucose within mammalian cells, where the ^{13}C-isotopes are shown in black. Dashed lines indicate more than one reaction between metabolites. **B,** Measured glucose uptake and lactate secretion in D492 and D492M. Total contribution of glucose to (**C**) lactate, (**D**) citrate, and (**E**) palmitate was measured after culturing of D492 and D492M with 1,2-^{13}C-glucose for 6 hours. Metabolites: f6p, fructose 6-phosphate; g3p, glyceraldehyde 3-phosphate; pyr, pyruvate; co2, carbon dioxide; accoa, acetyl CoA; oac, oxaloacetate; αkg, α-ketoglutarate; suc, succinate; fum, fumarate; mal, malate; asp, aspartate. Enzymes: CS, citrate synthase; ACLY, ATP-citrate lyase. Pathways: PPP, pentose phosphate pathway.

reductively (**Fig. 2A**). To discriminate oxidative and reductive TCA cycle carbon flow, D492 and D492M cells were cultured with isotopic glutamine analogs labeled at either the one or five positions. These isotopomers of glutamine allow the specific quantification of glutamine to citrate either reductively (1-^{13}C-glutamine) or oxidatively (5-^{13}C-glutamine; **Fig. 2A**). The 5-^{13}C-glutamine analog can additionally quantify the contribution of glutamine to fatty acids solely through reductive carboxylation (**Fig. 2A**). Importantly, intracellular glutamate label from a labeled glutamine source is a direct measurement of the contribution of glutamine to the glutamate pool. Compared with D492, the glutamate pool in the D492M cells was not as dependent on glutamine as observed from lower labeling incorporation from both 1- and 5-^{13}C glutamine (Supplementary Figs. S4A and S4B). This was supported by NMR measurements (Supplementary Fig. S1C) and is likely due to higher amounts of glutamate being derived from elsewhere (e.g., transamination reactions and protein catabolism) in D492M. As we were specifically interested in the contribution of glutamine to metabolites beyond glutamate (e.g., citrate), we accounted for the differences in glutamine-to-glutamate labeling by dividing the total contribution from glutamine to the target metabolites with the total contribution of glutamine to glutamate in each cell line. The results therefore represent the relative contribution of glutamate to metabolites.

Approximately 60% of the citrate pool was derived from glutamate in both D492 and D492M as shown by the isotope enrichment from 5-^{13}C glutamine (**Fig. 2C**). Enrichment of ^{13}C in citrate derived from 1-^{13}C glutamine was however higher in D492M (**Fig. 2C**), supporting increased reductive carboxylation. These results were mirrored in the 5-^{13}C-glutamine-dependent labeling profiles of palmitate, where there was a six-fold increase in reductive contribution of glutamine in D492M cells (**Fig. 2D**).

Metabolic re-routing following EMT affects redox metabolism in D492 cells

The reductive pathway of glutamine to citrate is typically activated in hypoxia or following changes in electron transport chain activity (41, 43, 44). Reductive glutaminolysis affects the redox status of cells through NADPH which serves as a cofactor for the reversible isocitrate dehydrogenase enzymes, IDH1 and IDH2 (45). Interestingly, we observed an increased NADPH/NADP ratio in D492M over D492

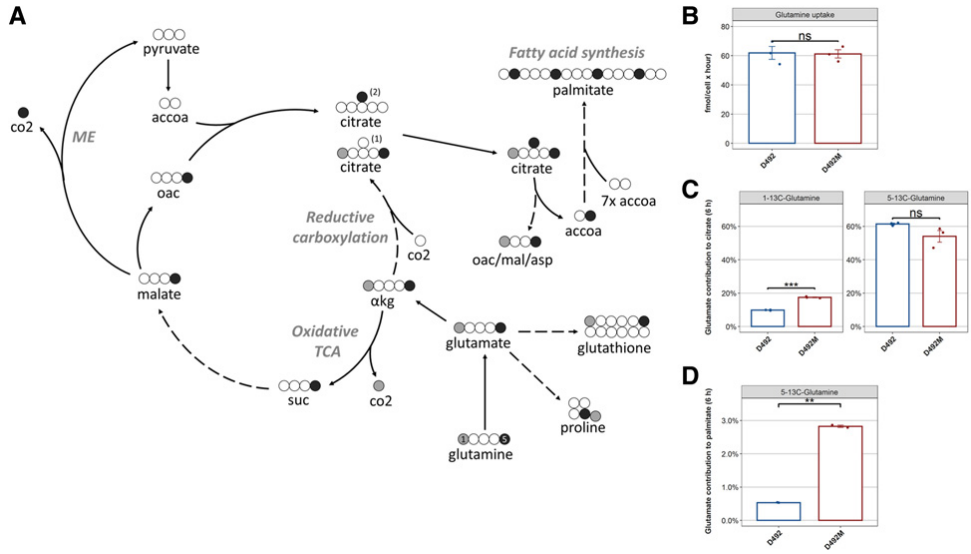


Figure 2.

Glutamine metabolism of D492 and D492M. **A,** Atom transition map of glutamine metabolism showing the different metabolic fates of 1–^{13}C glutamine (grey) and 5–^{13}C glutamine (black) within mammalian cells. Dashed lines indicate more than one reaction between metabolites. **B,** Measured glutamine uptake in D492 and D492M. **C,** Contribution of 1- and 5–^{13}C-glutamine to citrate in D492 and D492M cells after 6 hours in culture. The total contribution of the glutamine analogs to citrate was normalized to the different origins of glutamate in the cells (total contribution of glutamine to glutamate). **D,** Total contribution of 5–^{13}C-glutamine to palmitate after 6 hours in culture, normalized to glutamate origin. Metabolites: co2, carbon dioxide; accoa, acetyl CoA; oac, oxaloacetate; αkg, α-ketoglutarate; suc, succinate; mal, malate; asp, aspartate. Enzymes: ME, malic enzyme.

(**Fig. 3A**), which is not compatible with the PPP fluxes in the cells. However, even though reductive carboxylation is a NADPH-requiring process (**Fig. 3B**), studies have shown that its heightened activity may lead to increased cytosolic or mitochondrial NADPH levels based on the coordination of different isoforms of isocitrate dehydrogenase (45, 46). NADPH is known for its role in defenses against reactive oxygen species (ROS), where it is used to reduce the oxidized form of glutathione (GSSG) to generate the reduced form of GSH. Due to the observed alterations in glycolysis, glutamine metabolism, and NADPH/NADP$^+$ ratios, we hypothesized that this would be reflected in ROS generation within the cells. This in turn would translate to alterations in GSH metabolism.

Label incorporation from glutamine to GSH was a significantly lower following EMT in D492 cells as measured by UPLC-MS (**Fig. 3C**) and NMR (Supplementary Fig. S1C), further supporting EMT-associated alterations in redox metabolism. In addition, intracellular GSH concentration was 33% lower in D492M cells on average (**Fig. 3D**).

Proline has been shown to be important to maintain redox homeostasis, by recycling NADPH to NADP$^+$ (47), coupling it to NADPH-generating pathways. Proline is synthesized from glutamate in two reactions, both of which oxidize NADPH (**Fig. 3E**). We traced the proline synthesis from glutamate in D492 and D492M cells and observed a 2-fold increase in proline synthesis from glutamate following EMT (**Fig. 3F**). In summary, differences in the redox state of the D492 EMT model can be related to decreased glycolytic/PPP flux and concomitant changes to glutaminolysis, which is defined by decreased GSH synthesis, increased proline synthesis, and increased reliance on reductive carboxylation for citrate synthesis.

Mitochondrial isocitrate dehydrogenase 2 (IDH2) is essential for EMT-linked reductive glutamine metabolism

On the basis of the difference in reductive carboxylation and NADPH/NADP$^+$ ratio (**Figs. 2C** and **3A**, respectively) and recent literature (45, 48), we hypothesized that IDH would contribute to the discrimination between the D492 and D492M metabolic phenotypes (i.e., metabotypes) through glutamine consumption and influence redox balance.

Two isoforms of IDH are known to use NADP$^+$ as a cofactor: cytosolic IDH1 and mitochondrial IDH2. Quantification of IDH mRNA (by RNA sequencing) and protein (by MS proteomics) in D492 and D492M cells revealed increased IDH2 levels in D492M as compared with D492 (**Fig. 4A**). These findings were confirmed by Western blot analysis (Supplementary Fig. S5). No difference was observed in IDH1 levels (**Fig. 4A**). Using shRNA lentiviral transduction, we knocked down *IDH2* expression in both cell lines and investigated the metabolic and morphologic effects.

Following a significant reduction in *IDH2* expression levels, the mRNA levels of the cytosolic isoform IDH1 did not change in the D492M-IDH2 cell line (**Fig. 4B**). There were no differences observed in neither morphology nor growth rate upon *IDH2* knockdown in D492M (**Fig. 4C** and **D**). A significant reduction in reductive carboxylation activity was observed, as indicated by the 1–^{13}C-glutamine contribution to citrate and 5–^{13}C-glutamine contribution to palmitate

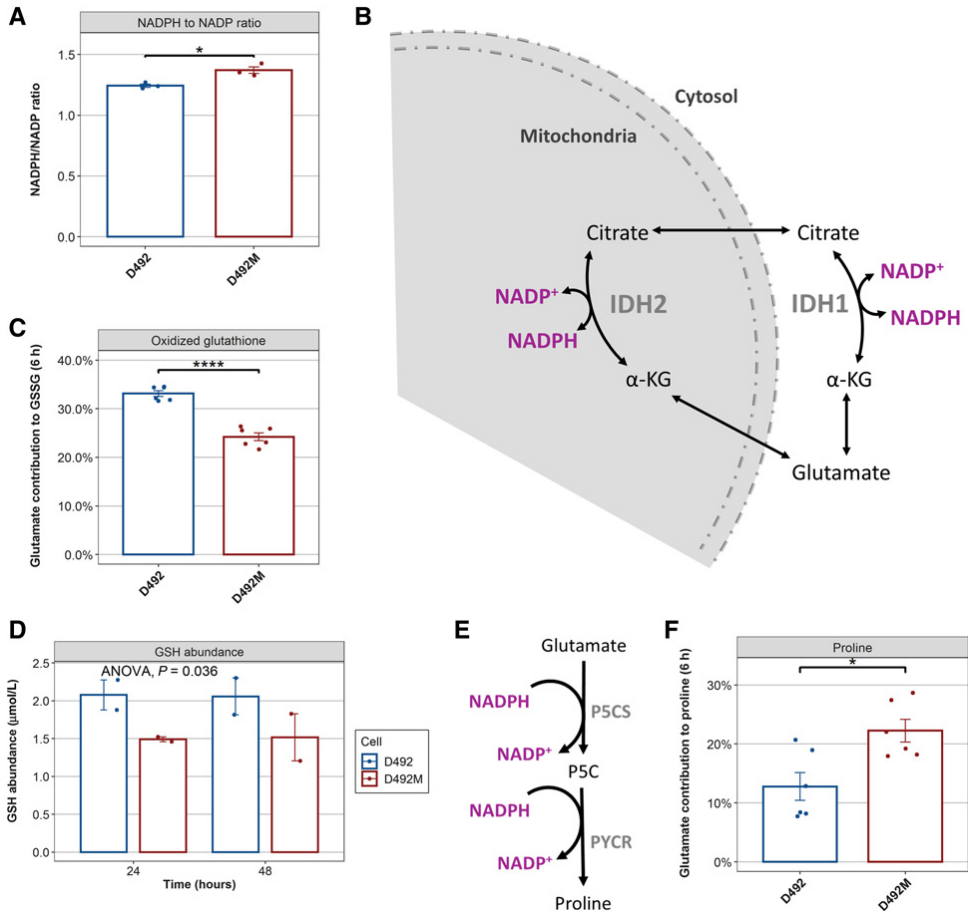


Figure 3.

Redox metabolism is altered following EMT of D492. **A,** NADPH-to-NADP ratio in D492 and D492M cells. **B,** A schematic showing the connection of mitochondrial and cytosolic reductive carboxylation, and $NADP^+$/NADPH balance. **C,** Total contribution of glutamate to oxidized glutathione (GSSG) in D492 and D492M cells after 6 hours in culture. **D,** Measured abundance of GSH in D492 and D492M after 24 and 48 hours in culture. A two-way ANOVA test revealed a significant difference in GSH levels between the cells, independent of time. **E,** The two reactions needed to convert glutamate into proline. Both reactions oxidize NADPH. P5CS: Delta-1-pyrroline-5-carboxylate synthase, PYCR, pyrroline-5-carboxylate reductase. **F,** Total contribution of glutamate to proline in D492 and D492M cells after 6 hours in culture. The results from **B** and **D** are from the combined analysis of 1- and 5-^{13}C-glutamine results, since both 13C carbons are detected in the proline and GSSG carbon skeletons.

(**Fig. 4E** and **F**, respectively). No significant difference was observed in the overall contribution of glutamine to citrate or glucose to citrate (**Fig. 4E**). The NADPH/$NADP^+$ ratio was lowered upon *IDH2* knockdown and proline synthesis from glutamate was reduced (**Fig. 4G** and **H**, respectively), suggesting that the redox homeostasis of these cells is coupled to IDH2-mediated reductive carboxylation. The glutamine-dependent synthesis of GSH, however, was not affected by the *IDH2* knockdown (**Fig. 4I**). In contrast to D492M cells, D492 cells increased the expression of the cytosolic isoform *IDH1* upon *IDH2* knockdown (Supplementary Fig. S6A). The upregulation of *IDH1* was accompanied by increased glutamate-to-lipid contribution via reductive carboxylation (Supplementary Fig. S6B) and decreased growth rate (Supplementary Fig. S6C). In addition, we observed a shift in NADPH-to-$NADP^+$ ratio, increased proline synthesis from glutamate, and reduced GSH synthesis in the D492-IDH2 knockdown cells compared with the wild-type D492 cells (Supplementary Figs. S6D–S6F). These results indicate that the epithelial phenotype, but not the mesenchymal phenotype of D492 cells, can compensate for the *IDH2* knockdown by increasing *IDH1* expression.

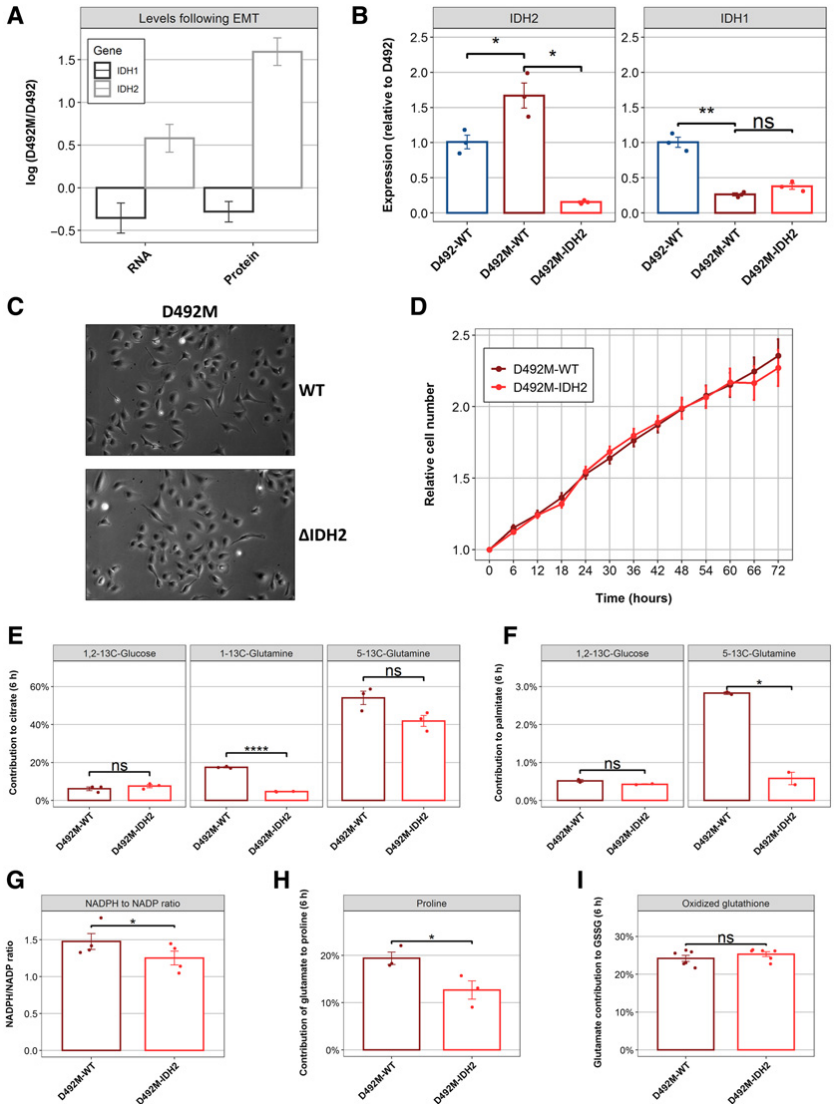


Figure 4.

IDH2 mediates reductive carboxylation activity and is coupled to redox metabolism of D492M. **A,** Relative differences in gene expression and protein translation of *IDH1* and *IDH2* in D492 and D492M. The results are displayed as log-fold D492M/D492 ratio of abundance of *IDH1/2* transcripts from RNA sequencing, and IDH1/2 protein from a proteomic analysis of D492 and D492M. Results are depicted as mean + standard error ($n = 3$). **B,** qPCR from D492, D492M and an *IDH2* -silenced D492M cell line showing the gene expression levels of *IDH2* and *IDH1*. **C,** Phase-contrast images of D492M-WT and D492M-IDH2 cells. **D,** Proliferation of D492M-WT and D492M-IDH2 cell lines (mean + SE, $n = 8$). **E,** Effect of *IDH2* silencing on the contribution of 1-^{13}C-glutamine, 5-^{13}C-glutamine and 1,2-^{13}C-glucose to citrate in D492M, where the former two were normalized to glutamate origin. **F,** Effect of *IDH2* silencing on the contribution of 5-^{13}C-glutamine (normalized to glutamate origin) and 1,2-^{13}C-glucose to palmitate in D492M. **G,** Effect of *IDH2* silencing on the NADPH/NADP$^+$ ratio in D492M. **H,** Effect of *IDH2* silencing on the contribution of 1-^{13}C-glutamine to proline (normalized to glutamate origin). **I,** Effect of *IDH2* silencing on the contribution of 1-^{13}C-glutamine and 5-^{13}C-glutamine to oxidized GSH (normalized to glutamate origin). **E, F, H,** and **I** are from cells cultured with ^{13}C-labeled carbon sources for 6 hours. Student two-tailed *t* test (with Benjamini–Hochberg adjustment for multiple comparisons) was used to estimate significance of results.

Alteration in GSH biosynthesis drives sensitivity to mTOR inhibition

The different metabotypes of the cell lines, characterized by different glycolytic rates, altered carbon source preference for TCA cycle activity and changes to the synthesis of proline and GSH are reminiscent of cancer stem cell metabotypes (49, 50). Because of GSH's role in drug resistance in various cancer cell types (51–53), we focused our attention on the differences in GSH metabolism between D492 and D492M cells (**Fig. 3**). We hypothesized that metabolic rerouting of glutamine-derived glutamate for GSH synthesis would affect drug sensitivity in these cells.

To identify drugs that are selectively affected by GSH concentrations within cells, we performed an integrated network analysis of (i) drug sensitivity profiles within the NCI-60 Human Tumor Cell Line database (29) and (ii) untargeted metabolomic analysis of NCI-60 cell lines from Ortmayr and colleagues (32). The network analysis revealed that 214 FDA-approved drugs in the NCI-60 database were grouped into eight intracorrelated drug modules (**Fig. 5A**), whose functional annotation showed that drugs grouped together according to their mechanism of action. The modules were comprised of DNA-damaging agents and cell-cycle arresting compounds (e.g., alkylating agents, nucleotide analogs, and paclitaxel, $n = 122$), hormones ($n = 7$), tyrosine kinase inhibitors targeting EGFR and ERBB2 (e.g., erlotinib and lapatinib, $n = 21$), mTOR and serine/threonine kinase inhibitors (e.g., everolimus, rapamycin, temsirolimus, $n = 8$), ALK/CDK inhibitors ($n = 17$), tyrosine kinases targeting VEGFR, PDGFR, and FGFR ($n = 7$), MAP kinase inhibitors $n = 10$), and nonspecific drugs ($n = 6$).

Interestingly, GSH levels were negatively correlated with the mTOR inhibitor module, represented in **Fig. 5A** as the size of the nodes. When the modules were represented collectively as a single unit using singular value decomposition, mTOR inhibitors were significantly correlated with both reduced and oxidized GSH (**Fig. 5B**). This was further supported by the Spearman correlation *P* value for mTOR inhibitors and intracellular GSH and GSSG levels (Supplementary Fig. S7). These results imply that high GSH availability is associated with low sensitivity to mTOR inhibitors and vice versa. Other conjugation metabolites (i.e., UDP-glucuronate and S-adenosylmethionine) did not display this type of relationship with mTOR inhibitors.

Following this we examined the effects of the mTOR inhibitor everolimus (strong negative correlation with GSH levels) and paclitaxel (no correlation with GSH levels) on D492 and D492M cells. D492M cells were more sensitive to both everolimus and paclitaxel than D492 (**Fig. 5C**). To establish a functional link between GSH abundance and mTOR inhibitors, we cotreated D492 and D492M cells with BSO, an inhibitor of the rate-limiting enzyme glutamate-cysteine ligase (GCL) in GSH synthesis (**Fig. 5D** and **E**), and either everolimus (**Fig. 5F**) or paclitaxel (**Fig. 5G**). The sensitivity of both D492 and D492M cells to everolimus was enhanced by cotreatment with BSO, whereas these effects were not observed when the cells were cotreated with paclitaxel and BSO. Together, these data suggest that GSH availability directly affects sensitivity to drugs that specifically affect the mTOR pathway.

Discussion

D492 and D492M cells represent only two of the numerous phenotypes within the spectrum of EMT (54). Herein, we have thoroughly characterized the central carbon metabolic activity of these cell types using ^{13}C isotope tracers, specialized metabolic assays, and shRNA-mediated knockdown of gene expression. Furthermore, we have evaluated the functional consequences of EMT-mediated differences in redox metabolism using both *in silico* and *in vitro* drug-sensitivity analyses.

IDH2 plays a key role in EMT in breast epithelium

The data presented here support previous findings where genome-scale metabolic models (12) and signal-network models constrained with cell type-specific transcriptomic data (53) predicted lower glycolytic activity of D492M cells compared with D492 cells. In addition, the data indicate that D492M cells increasingly rely on reductive carboxylation of glutamine to citrate via isocitrate dehydrogenase activity. Transcriptomic and proteomic data from D492 and D492M cells confirm that the predominant EMT-associated form of isocitrate dehydrogenase is the mitochondrial $NADP^+$-dependent IDH2, with D492M cells showing significantly higher levels on both the transcript and protein levels (**Fig. 4A**). Knockdown of *IDH2* by lentiviral shRNA induction caused a marked reduction of reductive carboxylation of glutamine to citrate in D492M cells but did not affect the overall contribution of either glucose or glutamine to citrate (**Fig. 4E**). These results were reflected in the label incorporation of 5–^{13}C-glutamine and 1,2–^{13}C-glucose to palmitate (**Fig. 4F**). The decrease in citrate labeling from 1–^{13}C-glutamine indicates that in these cells, reductive carboxylation primarily takes place within the mitochondria, as opposed to in the cytosol via IDH1 activity.

The knockdown of IDH2 and the subsequent re-routing of glutamine metabolism significantly diminished the ratio of NADPH to $NADP^+$ and the synthesis of proline from glutamate (**Fig. 4G** and **H**), consistent with the relationship of reductive carboxylation and proline synthesis to redox homeostasis (45, 55). However, there was no clear connection of IDH2 to GSH synthesis (**Fig. 4I**).

Our findings highlight the importance of IDH2 in the increased reductive carboxylation following EMT in breast epithelium. However, we cannot exclude the importance of IDH1 in this context. It is reasonable to assume that when reductive carboxylation takes place in the cytosol via IDH1, the resulting citrate is transported into the mitochondria, where IDH2 takes part in its ultimate oxidation as proposed previously (45). When IDH2 levels are diminished, the activity of this pathway would inevitably be halted. Nevertheless, our results demonstrate that *IDH2* knockdown significantly affects the reductive carboxylation of glutamine to citrate and ultimately fatty acids which establishes a functional role of IDH2 in this process. Thus, not only are the D492 and D492M cells different in their the overall lipid composition (56), but also the origin of lipid carbons.

Alterations in reductive carboxylation and redox metabolism follow EMT in breast

Our results show that glutamine-derived citrate is utilized for fatty acid synthesis in the D492 cell model, but the reliance on this pathway is enhanced following EMT. We show that there is a concurrent increase in $NADPH/NADP^+$ ratio and proline synthesis along with a decrease in GSH synthesis (**Fig. 3C–F**). It has previously been shown that anchorage-independent growth relies on increased reductive carboxylation and subsequent mitigation of mitochondrial ROS (45). The role of proline in anchorage-independent growth has also been demonstrated by showing that its degradation and cycling is higher in breast cancer cells grown in 3D culture than in 2D (55), which ultimately altered the $NADPH/NADP^+$ balance. Phang (57) hypothesized that in cancer cells, proline is directed away from protein synthesis and towards redox regulation, a pathway that proline has previously been shown to participate in within mammalian cells (58). More recently, it was shown that NADPH-dependent proline synthesis and reductive carboxylation act as alternative bins for electrons under

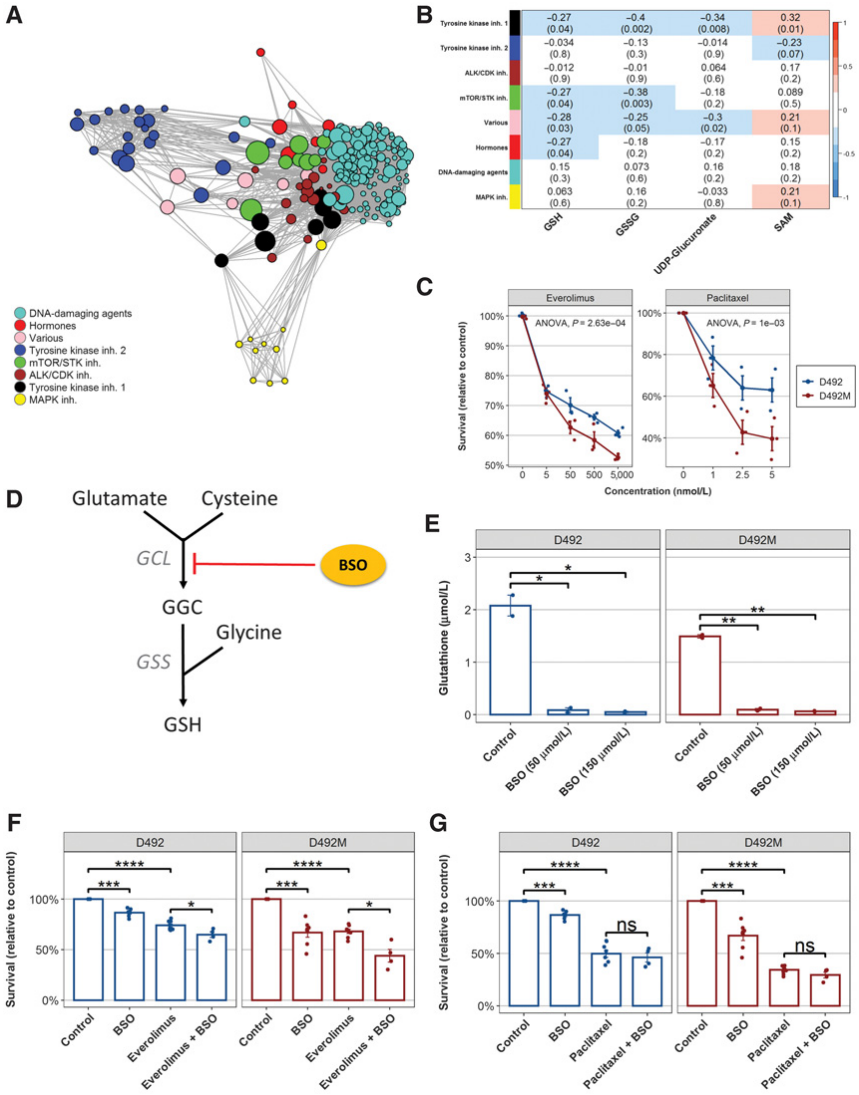


Figure 5.

GSH levels regulate sensitivity to mTOR inhibitors. **A,** Network analysis of NCI-60 cell lines treated with various FDA-approved drugs ($n = 214$) suggests the presence of 8 modules of intra-correlated drugs. The efficacy of each individual drug was correlated with GSH levels, represented by node size. **B,** The correlation of drug modules' eigenvalues to reduced GSH, oxidized glutathione (GSSG), UDP-glucuronate and S-adenosylmethionine (SAM). Upper numbers in table represent Pearson correlation coefficient and the lower numbers (in brackets) represent the correlation p-value. **C,** D492 and D492M cells treated with increasing concentrations of the mTOR inhibitor everolimus and taxane drug paclitaxel. Results are presented as mean + SE ($n \geq 4$). **D,** Mechanism of GSH synthesis inhibition by BSO. GGC, gamma-L-glutamyl-L-cysteine; GCL, glutamate-cysteine ligase; GSS, GSH synthetase. **E,** Effect of 24-hour BSO-treatment on GSH concentration in D492 and D492M cells. **F,** D492 and D492M cells treated with everolimus (0.005 µmol/L) with and without BSO (50 µmol/L). **G,** D492 and D492M cells treated with paclitaxel (0.005 µmol/L) with and without BSO (50 µmol/L). For **F** and **G**, the Y-axis represents percentage of viability compared with nontreated cells after 72 hours of treatment (mean + standard error, $n \geq 4$). Two-way ANOVA and Student two-tailed t test (with Benjamini–Hochberg adjustment for multiple comparisons) were used to measure significance of results in **C** and **E–G**, respectively.

hypoxic conditions when the electron transport chain is disabled. This way, electron transfer may continue functioning in cancer cells to maintain their viability (50). These observations fit well with the decreased oxygen consumption rate in D492M mesenchymal cells (12), their shift towards higher overall NADPH (cf. **Fig. 3B**), increased reductive carboxylation and higher proline synthesis. Furthermore, D492M cells display a concomitant decrease in GSH synthesis and intracellular abundance. Interestingly, Snail mediated EMT induction in MCF7 breast cancer cells has previously been shown to result in intracellular GSH reduction and elevation of ROS (59).

Diminished GSH abundance potentiates sensitivity to PI3K/akt/mTOR inhibitors

We have previously reported a reduction in oxidative phosphorylation following EMT in D492 cells (12). This causes a metabolic shift towards anaplerosis, an upregulation of pathways receiving otherwise ETC-directed electrons (i.e., proline synthesis and reductive carboxylation; ref. 50) and decreased glutamine-derived GSH synthesis. GSH is the most abundant nonprotein thiol in animal cells and it plays a crucial role in the conjugation phase of xenobiotic metabolism. This leads to increased water-solubility of foreign compounds (e.g., drugs or toxins) and their reduced efficacy (60, 61).

Because of the reduction in GSH synthesis and overall abundance in the mesenchymal D492M cells compared with epithelial D492 cells (**Fig. 3C** and **D**), we hypothesized that this would result in altered drug sensitivity of the mesenchymal phenotype. Integrated network analysis suggested that D492M cells are more sensitive to drugs that specifically target mTOR (**Fig. 5A** and **B**). Furthermore, the lack of a significant relationship between UDP-glucuronate and S-adenosylmethionine, compounds known to partake in the conjugation to xenobiotic compounds (9), suggests that mTOR inhibitors are specifically affected by GSH availability. We tested the efficacy of everolimus, a well-known mTOR inhibitor, and found that D492M cells were more sensitive than D492 cells (**Fig. 5C**). Furthermore, we showed that depleting intracellular GSH levels via BSO treatment (**Fig. 5E**) increased the sensitivity to everolimus (**Fig. 5F**). In contrast, BSO treatment did not affect sensitivity to paclitaxel, a microtubule stabilizer and mitotic inhibitor (**Fig. 5G**). These results indicate that GSH availability primarily affects the sensitivity to drugs that target the mTOR pathway.

In recent years, studies have shown a direct relationship between the mTOR signaling pathway and oxidative stress response (62, 63). Furthermore, mTOR signaling has been shown to be highly relevant in the EMT process and chemoresistance of ovarian cancer cells and melanoma (64, 65). Collectively, our results introduce a valuable mechanistic insight into the altered drug sensitivity following EMT in breast epithelium and support previous findings that GSH depletion in combination with mTOR inhibitors may specifically target the metastatic potential and/or stemness of cancer cells.

Conclusions

In summary, we have defined alterations in central carbon metabolism of a breast epithelial cell model of EMT using ^{13}C carbon tracing. We show that glutamine metabolism is re-routed towards reductive carboxylation to fuel fatty acid synthesis following EMT due to activity of the mitochondrial $NADP^+$-dependent isocitrate dehydrogenase (IDH2). This leads to decreased GSH production and disrupted redox homeostasis within the cells. Integrated network analysis of the NCI-60 Human Tumor Cell Line database revealed a negative correlation between intracellular GSH levels and sensitivity to mTOR inhibitors, and by depleting intracellular GSH levels in D492 and D492M cells, we sensitized them to mTOR pathway inhibitors. Our results highlight a potential metabolic weakness of low-GSH, EMT-derived cells that may be exploited in antimetastatic treatment.

Authors' Disclosures

S.T. Karvelsson reports grants from Göngum Saman and Icelandic Research Fund during the conduct of the study. A. Sigurdsson reports grants from Göngum Saman during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

Sigurdur Trausti Karvelsson: Conceptualization, data curation, software, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. **Arnar Sigurdsson:** Data curation, formal analysis, investigation, writing–review and editing. **Kotryna Seip:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **Maria Tunset Grinde:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **Qiong Wang:** Formal analysis, investigation, writing–review and editing. **Freyr Johannsson:** Formal analysis, investigation, methodology, writing–review and editing. **Gunhild Mari Mælandsmo:** Supervision, funding acquisition, writing–review and editing. **Siver Andreas Moestue:** Supervision, funding acquisition, investigation, writing–review and editing. **Ottar Rolfsson:** Conceptualization, data curation, formal analysis, supervision, funding acquisition, investigation, methodology, writing–original draft, writing–review and editing. **Skarphedinn Halldorsson:** Conceptualization, data curation, formal analysis, supervision, methodology, writing–original draft, writing–review and editing.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 4, 2020; revised March 16, 2021; accepted May 24, 2021; published first June 4, 2021.

References

1. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. Cancer Metastasis Rev 2016;35:645–54.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
3. Lee SY, Jeon HM, Ju MK, Jeong EK, Kim CH, Park HG, et al. Dlx-2 and glutaminase upregulate epithelial-mesenchymal transition and glycolytic switch. Oncotarget 2016;7:7925–39.
4. 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.
5. Koch K, Hartmann R, Schröter F, Suwala AK, Maciaczyk D, Krüger AC, et al. Reciprocal regulation of the cholinic phenotype and epithelialmesenchymal transition in glioblastoma cells. Oncotarget 2016;7:73414–31.
6. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong STC, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature 2015;527:472–6.
7. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature 2015;527:525–30.
8. Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial–mesenchymal plasticity in cancer metastasis. Nat Rev Cancer 2019;19:716–32.
9. Jancova P, Anzenbacher P, Anzenbacherova E. Phase II drug metabolizing enzymes. Biomed Pap 2010;154:103–16.
10. Jóhannsson F, Guðmundsson S, Paglia G, Guðmundsson S, Palsson B, Sigurjónsson ÓE, et al. Systems analysis of metabolism in platelet

concentrates during storage in platelet additive solution. Biochem J 2018;475:2225–40.
11. Grinde MT, et al. Biomarker discovery using NMR-based metabolomics of tissue. New York, NY: Humana; 2019. p. 243–62. doi:10.1007/978-1-4939-9690-2_15.
12. Halldorsson S, Rohatgi N, Magnusdottir 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.
13. Sigurdsson V, Hilmarsdottir B, Sigmundsdottir H, Fridriksdottir AJR, Ringnér M, Villadsen R, et al. Endothelial induced EMT in breast epithelial cells with stem cell properties. PLoS One 2011;6:e23833.
14. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9:676–82.
15. Itkonen HM, Gorad SS, Duveau DY, Martin SES, Barkovskaya A, Bathen TF, et al. Inhibition of O-GlcNAc transferase activity reprograms prostate cancer cell metabolism. Oncotarget 2016;7:12464–76.
16. Austdal M, Skråstad RB, Gundersen AS, Austgulen R, Iversen A-C, Bathen TF. Metabolomic biomarkers in serum and urine in women with preeclampsia. PLoS One 2014;9:e91923.
17. Eilers PHC. Parametric time warping. Anal Chem 2004;76:404–11.
18. Savorani F, Tomasi G, Engelsen SB. icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 2010;202:190–202.
19. Bettum IJ, Gorad SS, Barkovskaya A, Pettersen S, Moestue SA, Vasiliauskaite K, et al. Metabolic reprogramming supports the invasive phenotype in malignant melanoma. Cancer Lett 2015;366:71–83.
20. Kotronoulas A, Jónasdóttir HS, Sigurðardóttir RS, Halldórsson S, Haraldsson GG, Rolfsson Ó. Wound healing grafts: omega-3 fatty acid lipid content differentiates the lipid profiles of acellular Atlantic cod skin from traditional dermal substitutes. J Tissue Eng Regen Med 2020;14:441–51.
21. Millard P, Letisse F, Sokol S, Portais JC. IsoCor: correcting MS data in isotope labeling experiments. Bioinformatics 2012;28:1294–6.
22. Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting 13C metabolite labeling patterns from cells. Curr Opin Biotechnol 2015;34:189–201.
23. Lee WNP, Boros LG, Puigjaner J, Bassilian S, Lim S, Cascante M, et al. Mass isotopomer study of the nonoxidative pathways of the pentose cycle with [1,2-13C2]glucose. Am J Physiol 1998;274:E843–51.
24. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016;34:525–7.
25. Briem E, Budkova Z, Sigurdardottir AK, Hilmarsdottir B, Kricker J, Timp W, et al. MiR-203a is differentially expressed during branching morphogenesis and EMT in breast progenitor cells and is a repressor of peroxidasin. Mech Dev 2019;155:34–47.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
27. 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 2019;47:D442–50.
28. 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.
29. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6:813–23.
30. Luna A, Rajapakse VN, Sousa FG, Gao J, Schultz N, Varma S, et al. Rcellminer: exploring molecular profiles and drug response of the NCI-60 cell lines in R. Bioinformatics 2016;32:1272–4.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
32. Ortmayr K, Dubuis S, Zampieri M. Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. Nat Commun 2019;10:1–13.
33. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Sy 2006;1695.
34. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp 1991;21:1129–64.
35. R Core Team. R: A language and environment for statistical computing; 2020.
36. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
37. Kassambara A. ggpubr: 'ggplot2' based publication ready plots; 2020.
38. Metallo CM, Walther JL, Stephanopoulos G. Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. J Biotechnol 2009;144:167–74.
39. Watson JA, Lowenstein JM. Citrate and the conversion of carbohydrate into fat. Fatty acid synthesis by a combination of cytoplasm and mitochondria. J Biol Chem 1970;245:5993–6002.
40. Gaglio D, Metallo CM, Gameiro PA, Hiller K, Danna LS, Balestrieri C, et al. Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. Mol Syst Biol 2011;7:523.
41. Yang C, Ko B, Hensley CT, Jiang L, Wasti AT, Kim J, et al. Glutamine oxidation maintains the TCA cycle and cell survival during impaired mitochondrial pyruvate transport. Mol Cell 2014;56:414–24.
42. Yoo HC, Yu YC, Sung Y, Han JM. Glutamine reliance in cell metabolism. Exp Mol Med 2020;52:1496–516.
43. Mullen AR, Wheaton WW, Jin ES, Chen P-H, Sullivan LB, Cheng T, et al. Reductive carboxylation supports growth in tumour cells with defective mitochondria. Nature 2012;481:385–8.
44. Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, et al. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. Nature 2012;481:380–4.
45. Jiang L, Shestov AA, Swain P, Yang C, Parker SJ, Wang QA, et al. Reductive carboxylation supports redox homeostasis during anchorage-independent growth. Nature 2016;532:255–8.
46. Du J, Yanagida A, Knight K, Engel AL, Vo AH, Jankowski C, et al. Reductive carboxylation is a major metabolic pathway in the retinal pigment epithelium. Proc Natl Acad Sci U S A 2016;113:14710–5.
47. Liu W, Hancock CN, Fischer JW, Harman M, Phang JM. Proline biosynthesis augments tumor cell growth and aerobic glycolysis: Involvement of pyridine nucleotides. Sci Rep 2015;5:17206.
48. Smolková K, Dvořák A, Zelenka J, Vítek L, Ježek P. Reductive carboxylation and 2-hydroxyglutarate formation by wild-type IDH2 in breast carcinoma cells. Int J Biochem Cell Biol 2015;65:125–33.
49. Aguilar E, Marin de Mas I, Zodda E, Marin S, Morrish F, Selivanov V, et al. Metabolic reprogramming and dependencies associated with epithelial cancer stem cells independent of the epithelial-mesenchymal transition program. Stem Cells 2016;34:1163–76.
50. Liu M, Wang Y, Yang C, Ruan Y, Bai C, Chu Q, et al. Inhibiting both proline biosynthesis and lipogenesis synergistically suppresses tumor growth. J Exp Med 2020;217:e20191226.
51. Yang CH, Kuo ML, Chen JC, Chen YC. Arsenic trioxide sensitivity is associated with low level of glutathione in cancer cells. Br J Cancer 1999;81:796–9.
52. Rocha CRR, Garcia CC, Vieira DB, Quinet A, de Andrade-Lima LC, Munford V, et al. Glutathione depletion sensitizes cisplatin- and temozolomide-resistant glioma cells in vitro and in vivo. Cell Death Dis 2014;5:e1505.
53. Nunes SC, Serpa J. Glutathione in ovarian cancer: a double-edged sword. Int J Mol Sci 2018;19:1882.
54. Zhang Y, Weinberg RA. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. Front Med 2018;12:361–73.
55. Elia I, Broekaert D, Christen S, Boon R, Radaelli E, Orth MF, et al. Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. Nat Commun 2017;8:15267.
56. 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.
57. Phang JM. Proline metabolism in cell regulation and cancer biology: recent advances and hypotheses. Antioxid Redox Signal 2019;30:635–49.
58. Krishnan N, Dickman MB, Becker DF. Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. Free Radic Biol Med 2008;44:671–81.
59. Mezencev R, Matyunina LV, Jabbari N, McDonald JF. Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity. BMC Cancer 2016;16:236.
60. Zhang K, Mack P, Wong KP. Glutathione-related mechanisms in cellular resistance to anticancer drugs (Review). Int J Oncol 1998;12:871–82.
61. Forman HJ, Zhang H, Rinna AG.: Overview of its protective roles, measurement, and biosynthesis. Mol Aspects Med 2009;30:1–12.
62. Li J, Shin S, Sun Y, Yoon S-O, Li C, Zhang E, et al. MTORC1-driven tumor cells are highly sensitive to therapeutic targeting by antagonists of oxidative stress. Cancer Res 2016;76:4816–27.

63. Hambright HG, Meng P, Kumar AP, Ghosh R. Inhibition of PI3K/AKT/mTOR axis disrupts oxidative stress-mediated survival of melanoma cells. Oncotarget 2015;6:7195–208.
64. Deng J, Bai X, Feng X, Ni J, Beretov J, Graham P, et al. Inhibition of PI3K/Akt/mTOR signaling pathway alleviates ovarian cancer chemoresistance through reversing epithelial-mesenchymal transition and decreasing cancer stem cell marker expression. BMC Cancer 2019; 19:618.
65. Seip K, Fleten KG, Barkovskaya A, Nygaard V, Haugen MH, Engesæter BØ, et al. Fibroblast-induced switching to the mesenchymal-like phenotype and PI3K/mTOR signaling protects melanoma cells from BRAF inhibitors. Oncotarget 2016;7:19997–20015.

Paper II

npj | Systems Biology and Applications

www.nature.com/npjsba

ARTICLE OPEN

Check for updates

Argininosuccinate lyase is a metabolic vulnerability in breast development and cancer

Sigurdur Trausti Karvelsson[1], Qiong Wang[1], Bylgja Hilmarsdottir[2], Arnar Sigurdsson[3], Siver Andreas Moestue[4,5], Gunhild Mari Mælandsmo[2], Skarphedinn Halldorsson[1,6], Steinn Gudmundsson[1,7] and Ottar Rolfsson[1] ✉

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 ^{13}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[1]. 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[2,3], IDH1 involved in the tricarboxylic acid (TCA) cycle[4], ACLY in fatty acid synthesis[5], and PRODH from proline synthesis[6]. Definitive metabolic patterns that differentiate between invasive and non-invasive cancer cells however remain elusive[7].

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)[8–10]. Metabolic alterations are believed to be a hallmark of cancer and tumor progression[11] 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[12–14]. 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[15]. Furthermore, the predicted metabolic variability between patients has been utilized successfully for their prognosis[14]. These studies[14,15] 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[16]. Similar cell models previously used to study EMT include HMLE and the PMC42 EMT cell models[17–19].

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[16] mirrored results from a comprehensive analysis of EMT metabolism[20] and anchorage-independent growth[21]. 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

[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland. [2]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. [3]Department of Chemistry, Sekr. TC 2, Faculty II Mathematics and Natural Sciences, Technische Universität Berlin, Berlin, Germany. [4]Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway. [5]Department of Pharmacy, Nord University, Namsos, Norway. [6]Institute for Surgical Research, Vilhelm Magnus Laboratory, Oslo University Hospital, Oslo, Norway. [7]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. ✉email: ottarr@hi.is

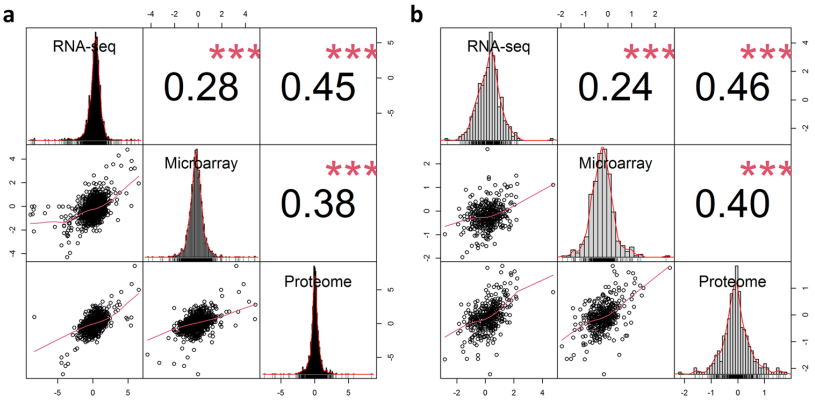


Fig. 1 Correlation of the log-fold differences in D492 and D492M of common gene identifiers between RNA sequencing, microarray and proteomic data. a Correlation of log-fold differences of all common gene identifiers ($n = 2271$). **b** Correlation of log-fold differences of common metabolic gene identifiers ($n = 395$). 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[22,23].

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 metabotypes 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 reconcilidate the predicted differences in metabolic phenotypes based on the different data types, we carried out enzyme inhibitor assays, 1-^{13}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[24]. 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[16] 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[25], 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-^{13}C-labeled glutamine experiments. This tracer is capable of quantifying the contribution of glutamine, one of two major carbon sources of D492 and D492M[16], 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 (c.f 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

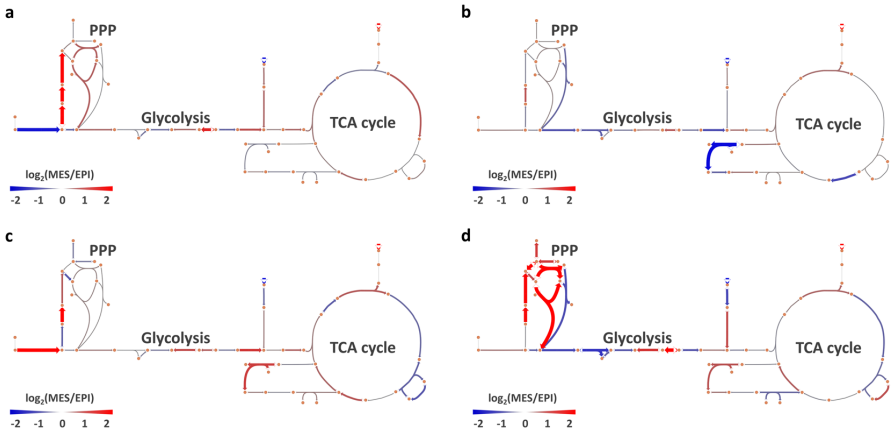


Fig. 2 Relative differences in reaction activity in EPI and MES models constrained in four different ways. a Only extracellular constraints, (**b**) Microarray, (**c**) RNA-seq and (**d**) Proteomic data. The pathways shown are glycolysis, TCA cycle, and pentose-phosphate pathway. Red represents higher 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-^{13}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[16] to identify reactions whose activity specifically requires alteration in order to switch from a 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[26]. 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[27]. 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[28]. *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[29] 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.

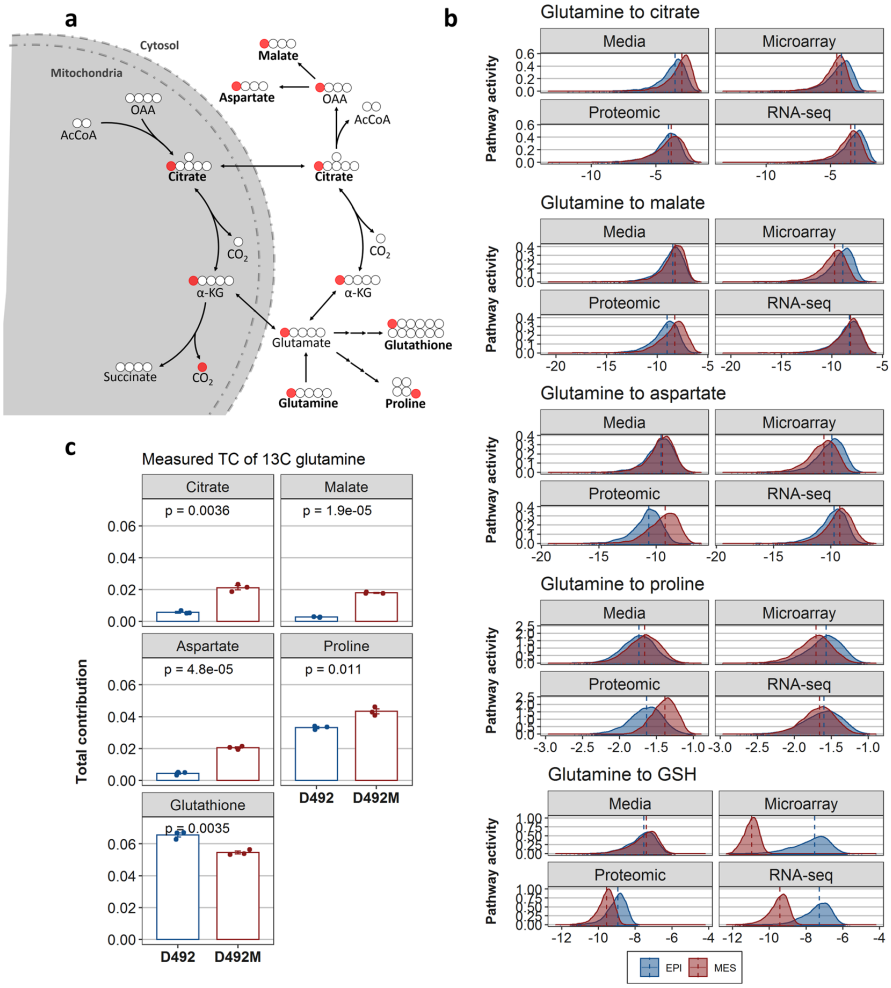


Fig. 3 Predicted and measured metabolic route activity in D492 and D492M cells. a A carbon tracing map of 1-^{13}C-labeled glutamine describing the flow and fate of labeled carbons in the glutamine carbon skeleton. Metabolites in bold are the end metabolites within the pathways we quantified. Red circles represent the ^{13}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 1-^{13}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-

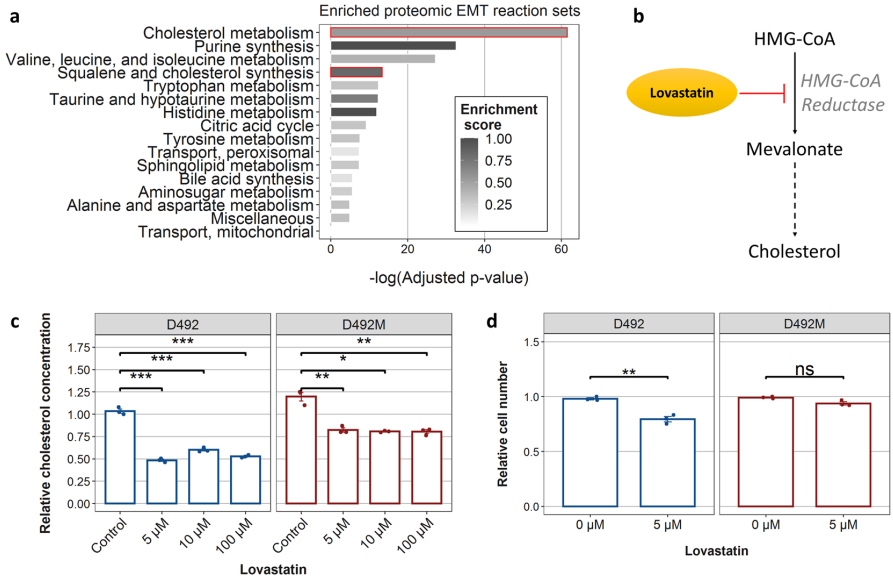


Fig. 4 Integrated analysis of all data-type based GSMMs reveals EMT-related differences in cholesterol metabolism. a Enrichment analysis of reactions that need alterations for the EPI to take on a MES flux phenotype in all proteomic GSMMs. This approach helps to identify reaction sets (i.e., families 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 μM concentration of lovastatin. Results in c) and d) 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.

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[30].

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[22,31–33]. 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[34].

The correlation of transcriptomic and proteomic data can be compartment-specific due to the different spatiotemporal nature of the molecules[35]. 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[36,37].

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[35]. 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 measured pathway activity using a 1-^{13}C-glutamine tracer (Fig. 3). Further analysis of the proteomic GSMMs was successful in proposing valid changes to D492

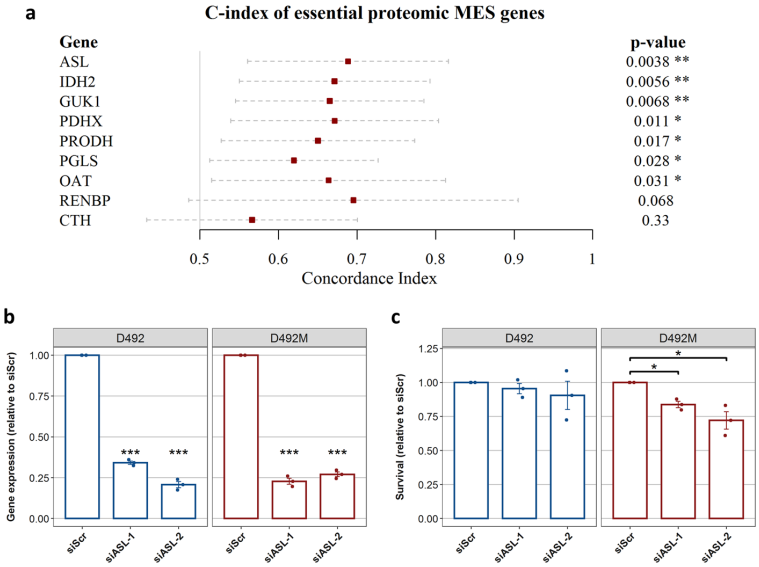


Fig. 5 Selection and knockdown of MES-essential genes. a Concordance index (C-index) of proteomic MES essential genes for overall survival prediction of patients with claudin-low breast tumors. Lines represent 95% confidence intervals and *p* values are from the calculation of the C-index using Noether's method[79]. 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.

		Univariate			Multivariable		
ER-status	Variables	HR	95% CI	*p* value	HR	95% CI	*p* 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[38] and the cholesterol-inhibiting drug statin has been shown to inhibit cancer invasion and metastasis[39–41]. 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[14,42]. A manual investigation of the GSMM 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[6,43–45]. For example, increased expression of the *IDH2* gene has been shown to be overexpressed in endometrial, prostate, testicular, and advanced colon cancer[46–48], 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[49].

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[14,15,50–52]. 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[6,53]. Downregulation of *ASL* has been shown to inhibit the growth of breast cancer tumors in vitro and in vivo[54]. Rabinovich et al.[55] 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[56]. Intercellular exchange of argininosuccinate between *ASL*- and *ASS1*-deficient cells, as demonstrated by Davidson et al.[57], 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[58]. The findings additionally support more studies that have shown that metabolic vulnerabilities of breast cancer lie within arginine metabolism[54,59,60].

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[14,61,62] and the relationship of cholesterol to EMT and metastasis has been discussed here above[38–41]. 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[63] 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% CO2 as previously described[19]. 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^{-10} 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[64]. The genome-scale breast tissue metabolic model from Halldorsson et al.[16] was used as a base model. Briefly, RNA sequencing data from both the D492 and D492M cell lines[19] was used to create a metabolic model specific for breast tissue. To achieve this, the human metabolic reconstruction Recon 2 was employed[65]. 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[66] 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-^{13}C-labeled glutamine (Cambridge Isotope Laboratories, Inc., MA, USA). After 6 h of culturing with the ^{13}C-labeled carbon source, cell metabolism was quenched using cold methanol and intracellular metabolites were extracted using ACN extraction[67]. Analyses were performed on a UPLC system as described in Rolfsson et al.[68]. Results were presented as the total contribution (TC) of carbon sources to measured metabolites[69]:

$$TC = \frac{\sum_{i=0}^{n} i \cdot m_i}{n} \quad (1)$$

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}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[70].

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 $m + k$, 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 µM 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™ 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[71] was downloaded from cBioPortal[72,73]. 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() function) of a gene of interest. The prognoses of the groups were then examined by calculating the concordance index (C-index)[28], 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_2 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 ug 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™ 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[74]. Primers spanning exon junctions were chosen to ensure specificity. Differences in relative expression were estimated using the $2^{\Delta\Delta Ct}$ method. The primer sequences used for quantifying the gene expression were: ASL-fwd 5'-GGAAGCTGTGTTTGAAGTGTCA-3', ASL-rev 5'-CCATGTTCTCTTGGTGAATCTG-3', ASS1-fwd 5'-CAGGAAAGGGGAACGATCAGGT-3', ASS1-rev 5'-CGTGTTGCTTTGCGTACTCCAT-3', GUK1-fwd 5'-CTTCATCGAGCATGCCGAGTTC-3', GUK1-rev 5'-GAACCTGTATGGCACGAGCAAG-3', ACTB-fwd 5'-CTTCCTGGGTGAGTGGAGACTG-3' and ACTB-rev 5'-GAGGGAAATGAGGGCAGGACTT-3'.

Analysis of clinical breast cancer data using iBreast2886

Proteomic breast cancer data were acquired from Tang et al.[29]. After removing identifers 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.[75]:

$$\hat{p} = \frac{r + 1}{n + 1} \quad (4)$$

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[29] 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[76]. To assess the effects of genes and confounding variables on patient survival, Cox-proportional hazard models were employed using the R-package survival[77].

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 + 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[78].

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.[29] 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.

Received: 29 January 2021; Accepted: 9 August 2021;
Published online: 17 September 2021

REFERENCES

1. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
2. 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).
3. 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).
4. 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).
5. 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).
6. Elia, I. et al. Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **8**, 15267 (2017).
7. Elia, I., Doglioni, G. & Fendt, S. M. Metabolic hallmarks of metastasis formation. *Trends Cell Biol.* **28**, 673–684 (2018).
8. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest* **119**, 1420–1428 (2009).
9. Ye, X. & Weinberg, R. A. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* **25**, 675–686 (2015).
10. 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).
11. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
12. Mardinoglu, A. & Nielsen, J. Editorial: the impact of systems medicine on human health and disease. *Front Physiol.* **7**, 552 (2016).
13. 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).
14. Yizhak, K. et al. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *eLife* **3**, e03641 (2014).
15. Jerby, L. et al. Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res.* **72**, 5712–5720 (2012).
16. Halldorsson, S. et al. Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett.* **396**, 117–129 (2017).
17. Tam, W. L. et al. Protein kinase C α is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell* **24**, 347–364 (2013).
18. Ackland, M. L. et al. Epidermal growth factor-induced epithelio-mesenchymal transition in human breast carcinoma. *Cells Lab. Investig.* **83**, 435–448 (2003).
19. Sigurdsson, V. et al. Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS ONE* **6**, e23833 (2011).
20. Shaul, Y. D. et al. Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **158**, 1094–1109 (2014).
21. Jiang, L. et al. Reductive carboxylation supports redox homeostasis during anchorage-independent growth. *Nature* **532**, 255–258 (2016).
22. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).
23. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl Acad. Sci. USA* **102**, 2685 (2005).
24. 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).
25. Schellenberger, J. & Palsson, B. O. Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* **284**, 5457–5461 (2009).
26. Istvan, E. S. & Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **292**, 1160–1164 (2001).
27. Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68–R68 (2010).
28. 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).
29. 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).
30. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
31. Edfors, F. et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12**, 883 (2016).
32. Darmanis, S. et al. Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* **14**, 380–389 (2016).
33. Gong, H. et al. Single-cell protein-mRNA correlation analysis enabled by multiplexed dual-analyte co-detection. *Sci. Rep.* **7**, 2776 (2017).
34. 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).
35. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
36. Lu, R. et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**, 358–362 (2009).
37. Cooper, G. M. *The Cell - A Molecular Approach* 2nd Edition. (Sunderland (MA): Sinauer Associates, 2000). citeulike-article-id:10266975.
38. Alikhani, N. et al. Mammary tumor growth and pulmonary metastasis are enhanced in a hyperlipidemic mouse model. *Oncogene* **32**, 961–967 (2013).
39. 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).
40. Nielsen, S. F., Nordestgaard, B. G. & Bojesen, S. E. Statin use and reduced cancer-related mortality. *N. Engl. J. Med.* **367**, 1792–1802 (2012).
41. Zhang, J. et al. Statins, autophagy and cancer metastasis. *Int. J. Biochem. Cell Biol.* **45**, 745–752 (2013).
42. Dai, Z. et al. Identification of cancer-associated metabolic vulnerabilities by modeling multi-objective optimality in metabolism. *Cell Commun. Signal.* **17**, 124 (2019).
43. Chen, B. et al. MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells. *BMC Cancer* **14**, 443 (2014).
44. 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).
45. 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).
46. 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).
47. Altenberg, B. & Greulich, K. O. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014–1020 (2004).
48. Lv, Q. et al. Altered expression levels of IDH2 are involved in the development of colon cancer. *Exp. Ther. Med.* **4**, 801–806 (2012).
49. 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. Bjornson, 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 beta-alanine 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 α 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 v.3.0. *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, Ó. 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 (13)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**, pl1 (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 Freyr Johannsson and Sarah McGarrity for valuable input regarding constraint-based modeling methodology, ^{13}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.

Paper III

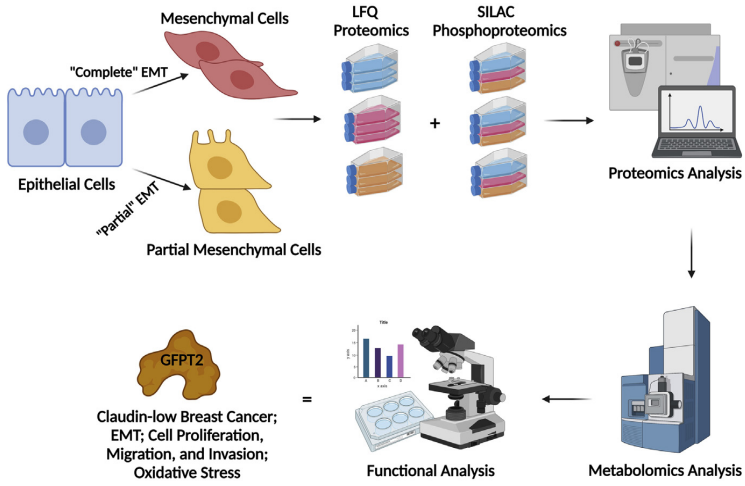
Glutamine-Fructose-6-Phosphate Transaminase 2 (GFPT2) Is Upregulated in Breast Epithelial–Mesenchymal Transition and Responds to Oxidative Stress

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

Correspondence
ottarr@hi.is

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.

Graphical Abstract



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

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[1], Sigurdur Trausti Karvelsson[1], Aristotelis Kotronoulas[1], Thorarinn Gudjonsson[2], Skarphedinn Halldorsson[1], and Ottar Rolfsson[1,*]

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 overexpresses 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_2S 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 [1]Center for Systems Biology, Biomedical Center, and [2]Stem Cell Research Unit, Biomedical Center, Department of Anatomy, Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavík, Iceland

*For correspondence: Ottar Rolfsson, ottarr@hi.is.

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

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

be triggered with growth factors such as TGF-β 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-κB and β-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_2 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 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), 1.4×10^{-6} 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 ^{13}C labeling experiment, the base medium was changed to "DMEM, no glucose, no glutamine, no phenol red" (Thermo, A1443001), and ^{13}C labeled 1,2-glucose, ^{13}C labeled 1-glutamine, or ^{13}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 μl 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 20,718*g* 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 μg of proteins in 150 μl 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 μg × 2), first overnight and then for another 6 h in a final volume of 200 μl. The resulting peptides were desalted using a C18 solid-phase extraction cartridge (Empore, Agilent technologies). Peptides were resuspended in 50 μl 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 μl/min onto an Acclaim PepMap100 nanoViper C18 trap column (100 μm inner-diameter, length: 2 cm; Thermo Scientific). After trap enrichment, peptides were eluted onto an EASY-Spray PepMap RSLC nanoViper, C18, particle size: 2 μm, pore size: 100 Å column (75 μm 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, gln → 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) LFQ: 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%. 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 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 LFQ 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-$^{13}C_6$ 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-$^{13}C_6$,$^{15}N_4$ hydrochloride (Arg +10 Da), L-lysine-$^{13}C_6$,$^{15}N_2$ 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 LFQ proteomics experiment. The same procedures as described in the LFQ 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 cOmplete mini EDTA-free protease inhibitors (Roche).

Protein Digestion (FASP Processing of Samples). Proteins were solubilized in 150 µl 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 µl iodoacetamide (IAA) at a final concentration of 50 mM for 15 min, filters were washed four times with 200 µl 8 M urea in Tris-HCl (100 mM, pH 8), then twice with 200 µl 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 µg × 2), first overnight and then for another 6 h in 200 µl, 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 µl 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 µm, inner-diameter: 4.6 × length: 150 mm, Ireland) with a guard column (XBridge, C18, particle size: 3.5 µm, inner-diameter: 4.6 × 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 µl) 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 µl each). Each peptide fraction was acidified immediately after elution from the column by adding 20 to 30 µl 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 µl, 1% NH_4OH or ammonia, followed with 200 µl acetonitrile. TiIMAC beads were equilibrated for 2 min with gentle mixing in 200 µl loading buffer consisting of 1 M glycolytic acid 80% acetonitrile and 5% trifluoroacetic acid (TFA). Dried samples were resuspended in 100 µl 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 µl loading buffer, three times with 200 µl of 80% acetonitrile-1% TFA, and finally with 200 µl of 10% acetonitrile-0.2% TFA. Phospho-peptides were eluted from beads three times using 80 µl of 1% ammonia, and gentle mixing with pH immediately lowered to 2 using 10% formic acid. Eluted phospopeptides were pooled, dried in speed vac at RT, and stored at −80 °C before LC-MS analysis.

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 μl of 1% formic acid. An aliquot (10 μl of total proteome; 15 μl of phospho-proteome) of each fraction was loaded at 10 μl/min onto a trap column (inner diameter: 100 μm × length: 2 cm, PepMap nanoViper C18 column, particle size: 5 μm, 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 μm × length: 50 cm, PepMap RSLC C18 column, particle size: 2 μm, 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-Orbitap 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 → 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 phosphoproteomic 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_2 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, β-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×) with 5% CO_2 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 μm 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 μl 1:10 diluted Matrigel for 20 to 30 min at 37 °C. Next, 300 μl of cell suspension was added on top of the filter units. Then, 500 μl 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_2 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×, 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 ^{13}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 ^{13}C 1,2-glucose or ^{13}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 ^{13}C labeling experiments, data were analyzed in ISOCORE, and we normalized the mean enrichment of ^{13}C in UDP-N-acetylglucosamine (UDP-GlcNAc) to the total amount of UDP-GlcNAc and presented it as relative ^{13}C incorporation. Metabolomic data were normalized to protein levels.

Hydrogen Peroxide (H_2O_2) 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 μM hydrogen peroxide (H_2O_2, 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_2O_2 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 μl 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) LINCS 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.

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 corresponded 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. 1*C*). 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. 1*C*). 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. 1*D* 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. 1*E*). 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

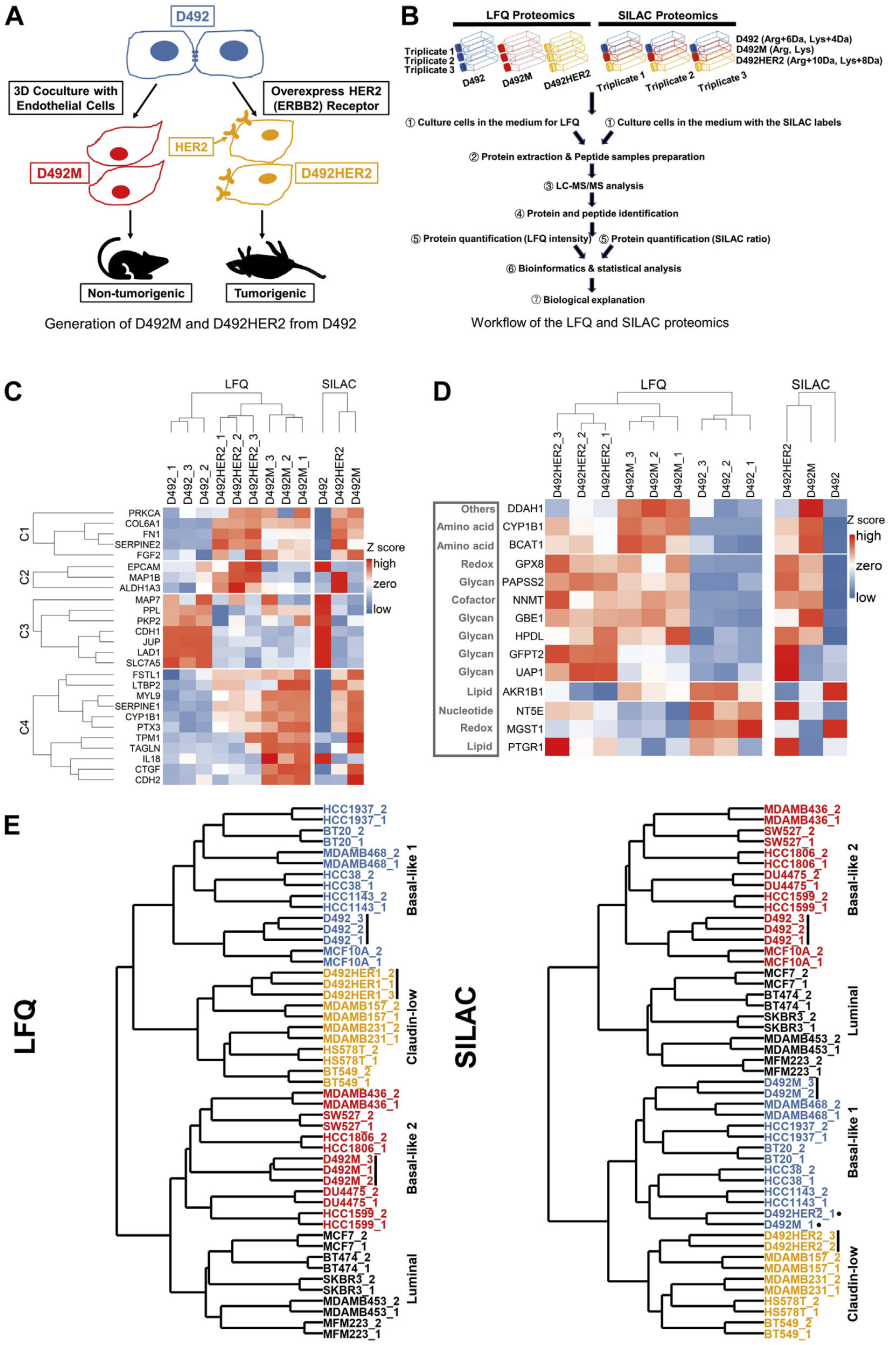


FIG. 1. **Overview of the study and the D492 EMT cell model.** *A*, the process of generating D492M and D492HER2 from D492. The D492 cells were cocultured with endothelial cells (BRENCs or HUVECs) to generate spindle colonies that were subcultured to generate a new cell line D492M. The D492M cells are 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

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. 4*A* and supplemental Fig. S5, *A–C*). The biggest difference in RNA expression was detected in GFPT2, consistent with the proteomic analysis (Fig. 4*B*). GFPT2 was not upregulated in D492 transfected with empty vectors (D492DEE) compared with D492HER2 (Fig. 4*C*). 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 maker, 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. S7*A*) as well as invasion in D492HER2 (Fig. 4, *M* and *N* and supplemental Fig. S7*B*). Decreasing trends were observed for migration after *GFPT2* knockdown in the three cell lines (Fig. 4*O* and supplemental Fig. S7*C*). No changes were observed in cell morphology (supplemental Fig. S6*G*). 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. 6*A*). 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. 6*B* 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}C isotopologue label enrichment in UDP-GlcNAc from cells grown in media containing 1,2-^{13}C glucose, 1-^{13}C glutamine, or 5-^{13}C glutamine. In context with the relative amount of UDP-GlcNAc, ^{13}C enrichment in UDP-GlcNAc from 1,2-^{13}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. 6*C* and supplemental Fig. S8, *D–F*). Little or no enrichment was observed in UDP-GlcNAc from 1-^{13}C glutamine. The m + 1 isotopologue in ^{13}C enrichment from the 5-^{13}C-glutamine was decreased in D492M compared with D492 and D492HER2 (supplemental Fig. S8*F*). Knockdown of *GFPT2* resulted in a clear decrease in the intracellular levels of UDP-GlcNAc in both D492M and D492HER2 (Fig. 6*D*), 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. 6*E*). We similarly noted a decrease in the intracellular cystathionine levels (Fig. 6*F*). 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. 7*A*). 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.

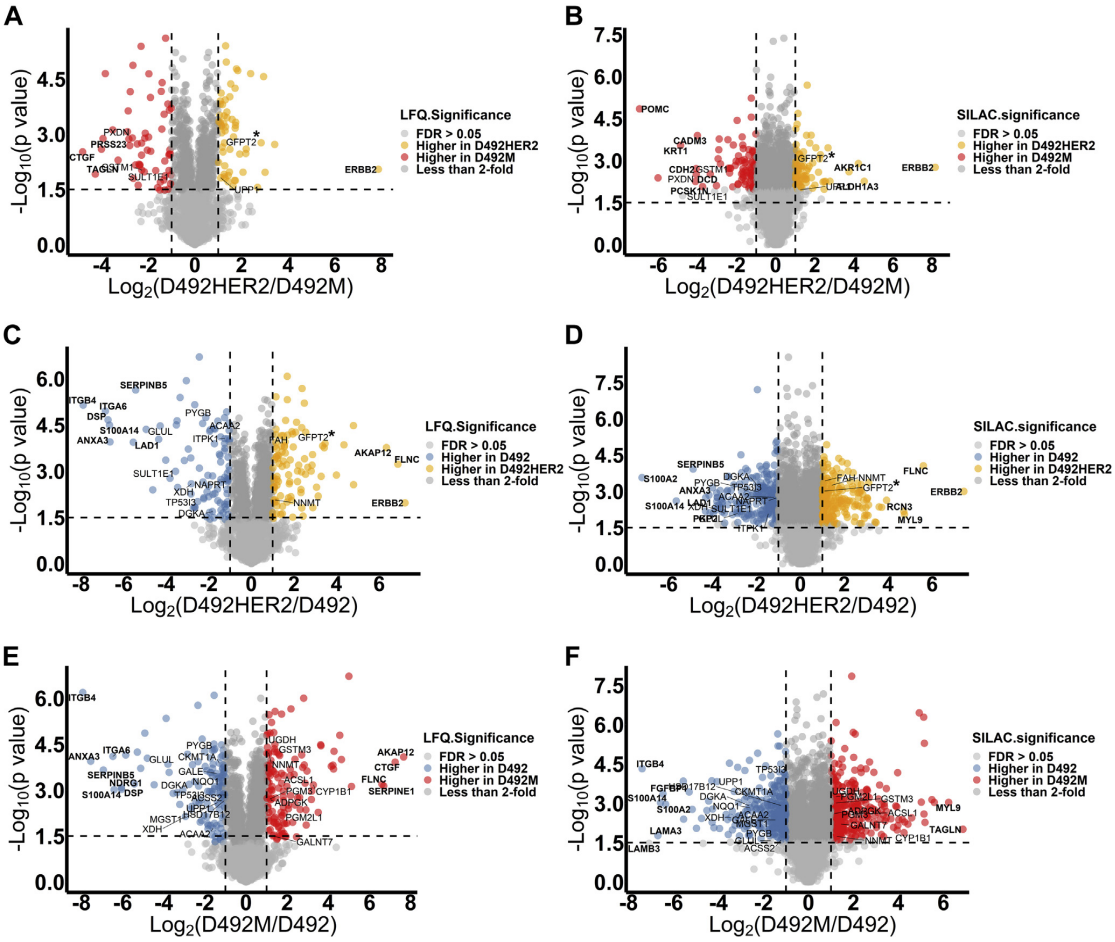


FIG. 2. **LFQ and SILAC proteomic data plotting.** Statistical analysis of the LFQ and SILAC expressions of proteome in two different cell lines: D492HER2 *versus* D492M (*A* and *B*), D492HER2 *versus* D492 (*C* and *D*), and D492M *versus* D492 (*E* and *F*). Proteins with FDR less than 0.05 and fold 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 $-\text{Log}_{10}$(*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 H_2O_2 treatment. We treated MDA-MB-231 cells with H_2O_2 and observed an increase of GFPT2 RNA expression (Fig. 7*B*) 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. 7*E*), confirming that GFPT2 expression reacts to GSH. The D492 cells possessed higher amounts of GSH than D492M and D492HER2 (Fig. 7*F*),

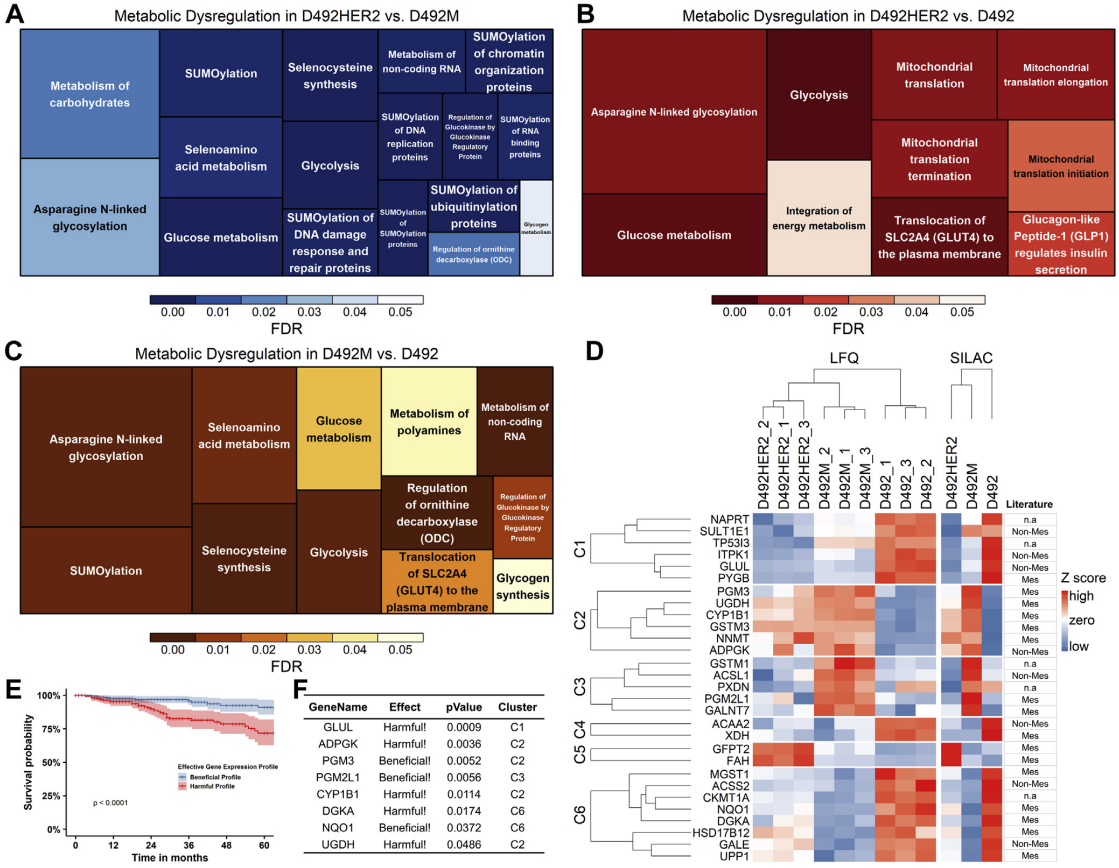


FIG. 3. **Dysregulated metabolic pathways in two mesenchymal transitions and the metabolic targets identified in this study.** *A-C,* Reactome metabolic pathways were differently enriched in D492HER2 *versus* D492M (*A*), D492HER2 *versus* D492, (*B*) and D492M *versus* D492 (*C*). Proteins involved in each Reactome metabolic pathway were plotted in supplemetal 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). *D*, 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 transsulfuration pathway and contributes to hydrogen sulfide (H_2S) and GSH synthesis. Considering that the levels of cystathionine dropped following *GFPT2* knockdown (Fig. 6*F*) 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_2S homeostasis to counteract oxidative stress. SQOR catalyzes the oxidation of H_2S 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-κB (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)	*GFPT2*	Carbohydrate metabolism	1.558	NF-KB; SIRT6; BMP-2	Shaul *et al.*, 2014 (44); Simpson *et al.*, 2012 (50); Szymura *et al.*, 2019 (53); Taparra *et al.*, 2019 (96); Zhang *et al.*, 2018 (55); Zhou *et al.*, 2019 (22)
Q16831	Uridine phosphorylase 1	*UPP1*	Pyrimidine metabolism	1.167	NF-Kb; Oct3/4	Guan *et al.*, 2019 (97); Wehbe *et al.*, 2012 (98)
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	Peroxidasin homolog	*PXDN*	Oxidoreductases	−3.825	Snail 1; Nrf2	Briem *et al.*, 2019 (16); Sitole and Mavri-Damelin, 2018 (99)
				Log2(D492HER2/D492)		
A0A0S2Z4X9	Glutamine-fructose-6-phosphate transaminase 2 isoform 1 (Fragment)	*GFPT2*	Carbohydrate metabolism	1.827	NF-KB; SIRT6; BMP-2	Shaul *et al.*, 2014 (44); Simpson *et al.*, 2012 (50); Szymura *et al.*, 2019 (53); Taparra *et al.*, 2019 (96); Zhang *et al.*, 2018 (55); Zhou *et al.*, 2019 (22)
Q6FH49	NNMT protein	*NNMT*	Nicotinate and nicotinamide metabolism	1.275	Stat3	Eckert *et al.*, 2019 (100); Shaul *et al.*, 2014 (44)
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 (101)
A0A024R6H3	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 (102)
A0A0B4J2A4	3-ketoacyl-CoA thiolase, mitochondrial	*ACAA2*	Lipid metabolism	−1.737	PPARα; HNF4α	n.a
A0A024RB23	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 (55)
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 (103); Reka *et al.*, 2014 (104)
Q53X91	Sulfotransferase (Fragment)	*SULT1E1*	Steroid hormone biosynthesis	−3.010	Nrf2	n.a
A8YXX4	Glutamine synthetase	*GLUL*	Carbohydrate metabolism	−3.582	ATF4	n.a
				Log2(D492M/D492)		
Q53TK1	Cytochrome P450, family 1, subfamily B, polypeptide 1, isoform CRA_a	*CYP1B1*	Lipid metabolism	3.733	SP1	Kwon *et al.*, 2016 (105); Shaul *et al.*, 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	*ACSL1*	Lipid metabolism	1.712	SP1	Sánchez-Martínez *et al.*, 2015 (106)
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 (107)
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 (100); Shaul *et al.*, 2014 (44)
Q9BRR6-2	Isoform 2 of ADP-dependent glucokinase	*ADPGK*	Glycolysis/Gluconeogenesis	1.173	n.a	Lee *et al.*, 2016 (108); Song *et al.*, 2018 (109)
O60701	UDP-glucose 6-dehydrogenase	*UGDH*	Carbohydrate metabolism	1.025	SP1	Tang *et al.*, 2016 (110); Vergara *et al.*, 2015 (111)
H0UIA1	Acyl-CoA synthetase short-chain family member 2, isoform CRA_c	*ACSS2*	Carbohydrate metabolism	−1.125	SREBF1/2; HIF; TFEB	Sun *et al.*, 2017 (112)
Q53GQ0	Very-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 (103); Reka *et al.*, 2014 (104)
Q14376	UDP-glucose 4-epimerase	*GALE*	Galactose metabolism	−1.458	n.a	n.a
Q6LET6	MGST1 protein (Fragment)	*MGST1*	Glutathione metabolism	−1.522	n.a	Fischer *et al.*, 2015 (113); Shaul *et al.*, 2014 (44)
P11216	Glycogen phosphorylase, brain form	*PYGB*	Starch and sucrose metabolism	−1.611	n.a	Zhang *et al.*, 2018 (55)
P12532	Creatine kinase U-type, mitochondrial	*CKMT1A*	Arginine and proline metabolism	−1.686	LncRNA n335586&miR-924; EVI1&RUNX1	Tanaka and Ogishima, 2015 (114)
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 (113); Yang *et al.*, 2017 (115)
Q16831	Uridine phosphorylase 1	*UPP1*	Pyrimidine metabolism	−2.303	NF-Kb; Oct3/4	Guan *et al.*, 2019 (97); Wehbe *et al.*, 2012 (98)
A0A024RB23	Diacylglycerol kinase	*DGKA*	Lipid metabolism	−2.730	PPARγ; Stat5; AP2, Ets1,SP1	n.a
A8YXX4	Glutamine synthetase	*GLUL*	Carbohydrate metabolism	−2.741	ATF4	n.a
P47989	Xanthine dehydrogenase/oxidase	*XDH*	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.

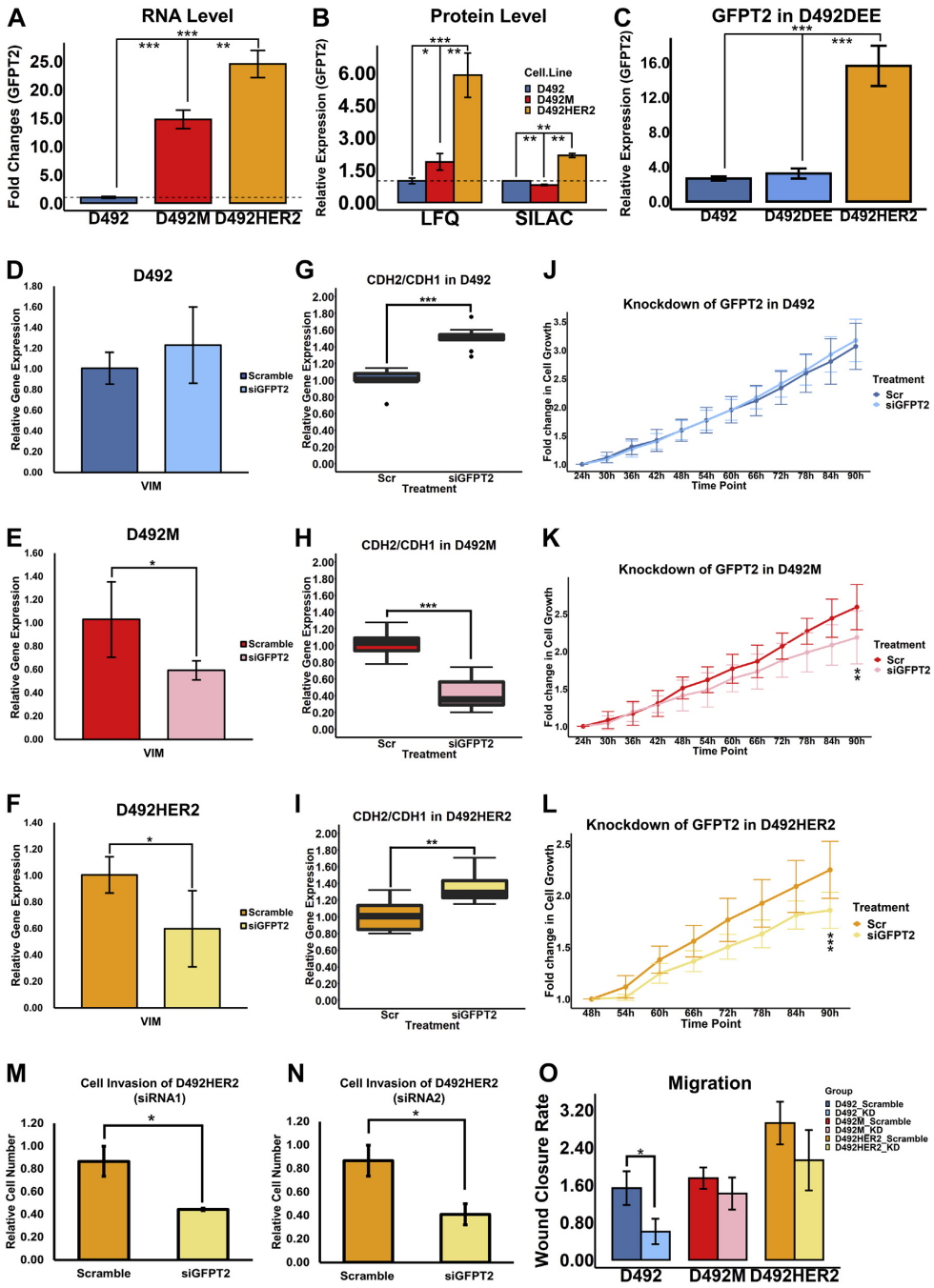


FIG. 4. **Expression and functions of GFPT2 in the D492 EMT model.** *A*, GFPT2 showed the highest expression in D492HER2 while lowest in D492 on the RNA level; *B*, The protein expression of GFPT2 in the three cell lines suggested highest expression of GFPT2 in D492HER2 confirmed by both LFQ (*left*) 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

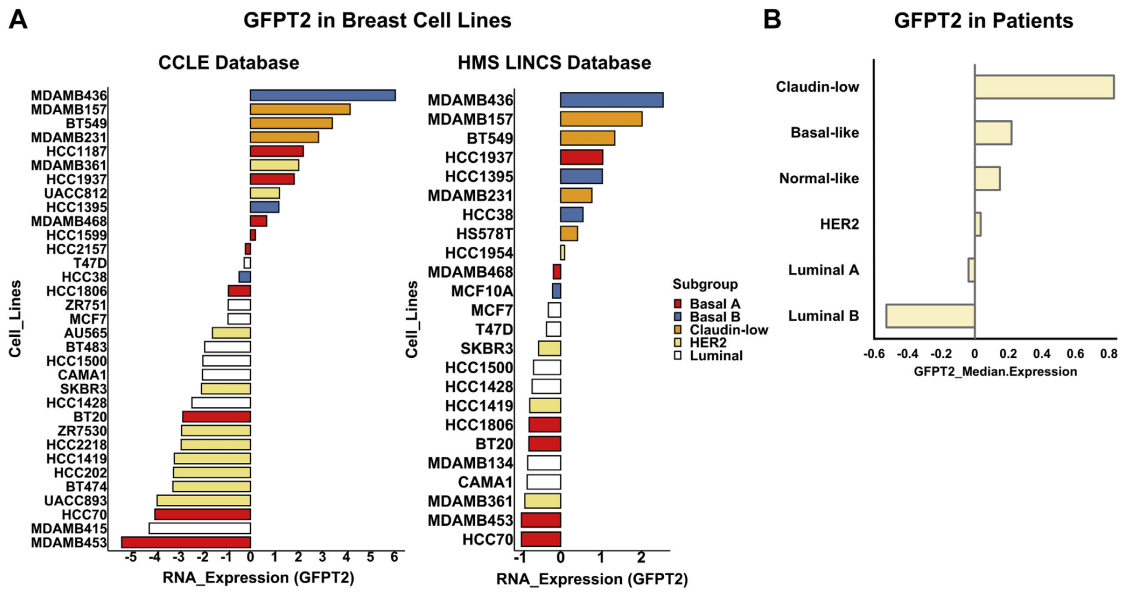


FIG. 5. **GFPT2 is higher in basal and claudin-low breast cell lines, and same trend is shown in breast cancer patients.** *A*, based on data from an open-source database – CCLE (*left*) (39), GFPT2 was higher in basal and claudin-low cell lines while lowly expressed in HER2-positive and luminal cell 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-β, IGF, TNF, and EGF signaling (Fig. 8*A*), 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. 8*D*), 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-β and PKCα substrates (Fig. 8, *E* and *F*). siRNA-mediated knockdown of *GSK3-β* 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. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

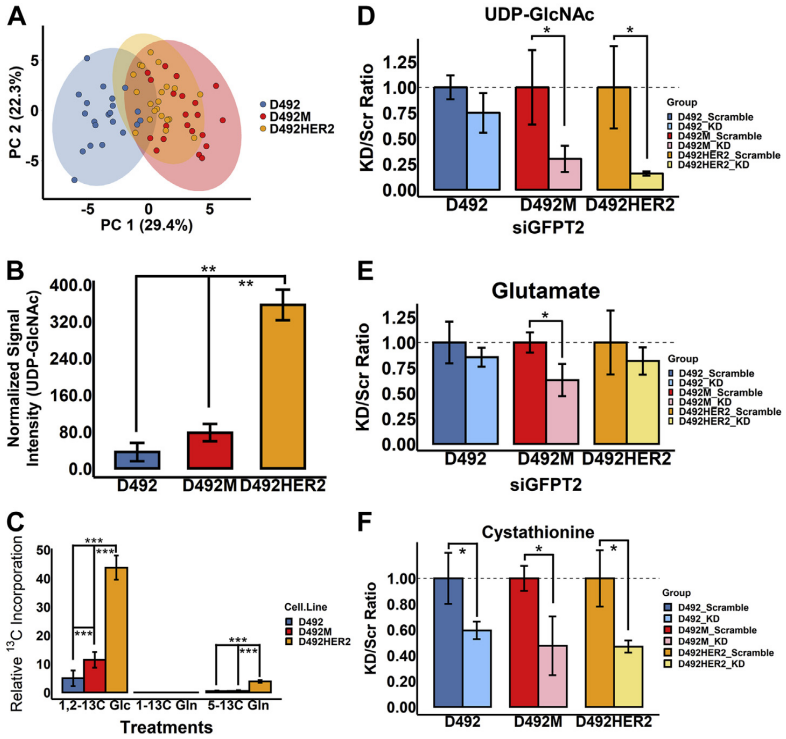


FIG. 6. **Metabolomic analyses of the D492 EMT cell model.** *A*, the metabolome in D492M and D492HER2 was similar compared with that in D492. *B*, UDP-GlcNAc was expressed higher in D492HER2 compared to the other cell types. *C*, carbon incorporation into UDP-GlcNAc after 6 h' culture. Carbons from 1,2-^{13}C glucose (Glc) were highly incorporated into UDP-GlcNAc, compared to 5-^{13}C glutamine (Gln) and 1-^{13}C Gln in all three cell lines. No carbon incorporation from 1-^{13}C Gln in all three cell lines. Higher rates of carbon incorporation into UDP-GlcNAc from both 1,2-^{13}C Glc and 5-^{13}C Gln were observed in D492HER2, compared to D492 and D492M. A higher rate of carbon incorporation into UDP-GlcNAc from 1,2-^{13}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 mammoplasty 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.

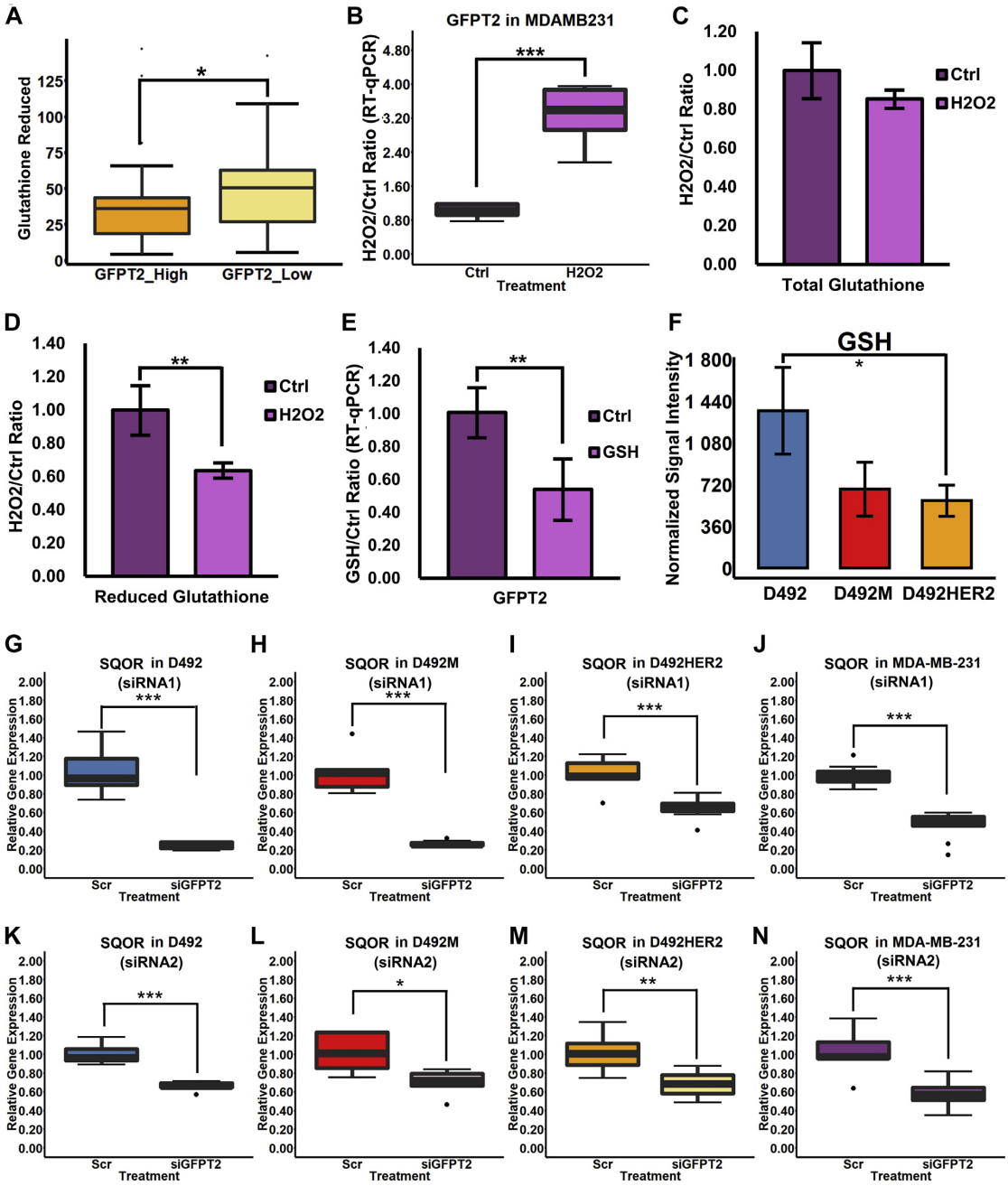


FIG. 7. **GFPT2 is a marker for cellular stress.** *A*, gene–metabolite correlation analysis of the NCI60 cancer cell line panel indicated a negative correlation between GFPT2 and GSH. *B*, GFPT2 RNA expression was significantly upregulated with 2 μM H_2O_2 treatment in MDA-MB-231. *C*, the total glutathione level did not change after H_2O_2 treatment. *D*, the GSH level was significantly decreased by the H_2O_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*). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

A

Protein ID	Protein Name	Gene Name	Log2(D492HER2/ D492)	-Log(p.value)	FDR
P01137	Transforming growth factor beta-1	TGFB1	5.208	3.669	0.000
P22692	Insulin-like growth factor-binding protein 4	IGFBP4	4.723	5.901	0.000
Q8WX77	Insulin-like growth factor-binding protein-like 1	IGFBPL1	3.821	2.760	0.002
O95407	Tumor necrosis factor receptor superfamily member 6B	TNFRSF6B	3.415	3.705	0.000
Q15582	Transforming growth factor-beta-induced protein ig-h3	TGFBI	2.644	4.449	0.000
Q9NP84	Tumor necrosis factor receptor superfamily member 12A	TNFRSF12A	1.902	1.144	0.037
P18065	Insulin-like growth factor-binding protein 2	IGFBP2	1.719	2.678	0.002
P61812	Transforming growth factor beta-2	TGFB2	1.684	1.564	0.014
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	1.080	2.622	0.003
C9JD84	Latent-transforming growth factor beta-binding protein 1	LTBP1	-1.770	1.931	0.006
Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2	EPS8L2	-1.896	2.188	0.005
P00533	Epidermal growth factor receptor	EGFR	-2.790	3.298	0.001
P08581	Hepatocyte growth factor receptor	MET	-3.450	4.062	0.000
Q14512	Fibroblast growth factor-binding protein 1	FGFBP1	-5.385	7.001	0.000

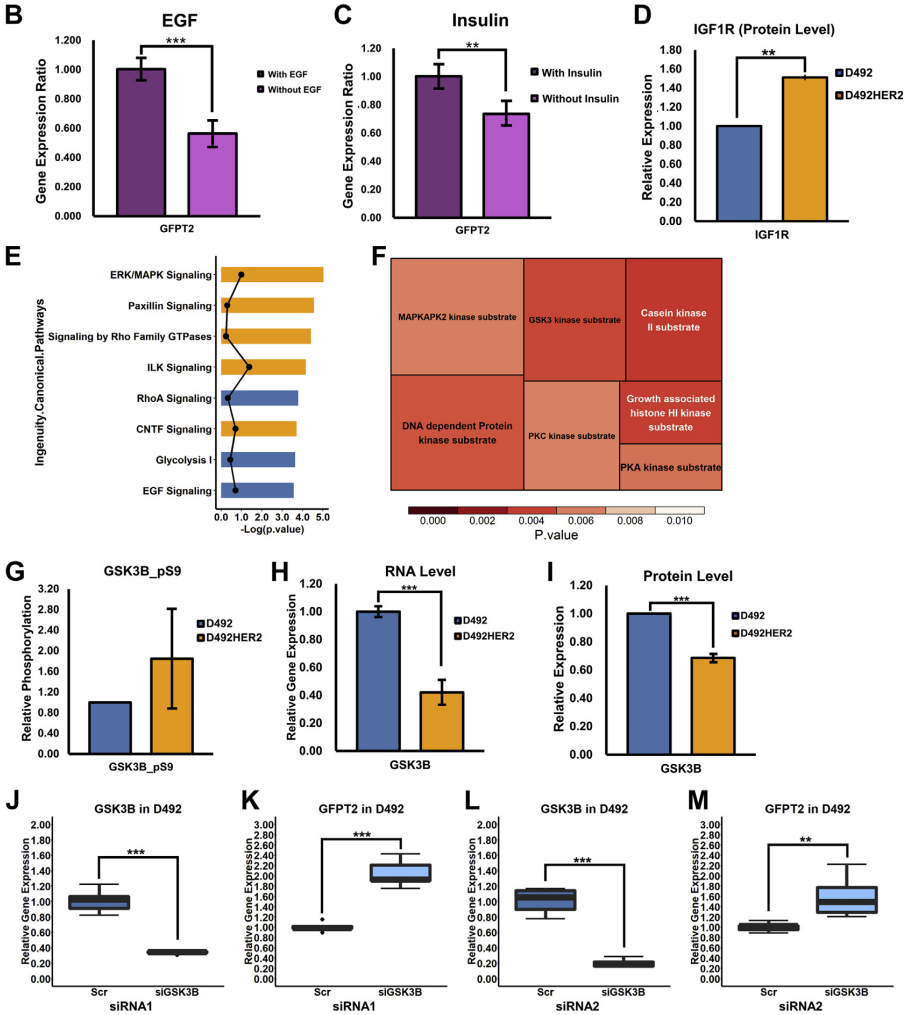


FIG. 8. **Signaling regulation of GFPT2.** *A*, secretome of D492HER2 and D492 revealed a list of growth factors that secreted differently between these two cell lines (FDR <0.05, Fold change ≥ 2). *B* and *C*, to test the effects of growth factors on GFPT2, we adapted the MDA-MB-231 cells with the FBS-free 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

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, ^{13}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). ^{13}C enrichment from the 5-^{13}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_2O_2 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-β was highly phosphorylated at position serine 9, which inhibits GSK3-β activation in D492HER2 compared with D492. *H* and *I*, RNA (*H*) and protein (*I*) expression of GSK3-β in D492HER2 *versus* D492 indicated the higher abundance of GSK3-β in D492. *J–M*, knockdown of *GSK3-β* in D492 increased GFPT2 RNA expression. *J*, knockdown efficiency for *GSK3-β* with the first siRNA. *K*, GFPT2 RNA expression after knockdown of *GSK3-β* in D492 with the first siRNA. *L*, knockdown efficiency for *GSK3-β* with the second siRNA. *M*, GFPT2 RNA expression after knockdown of *GSK3-β* in D492 with the second siRNA. $^{**}p < 0.01$; $^{***}p < 0.001$.

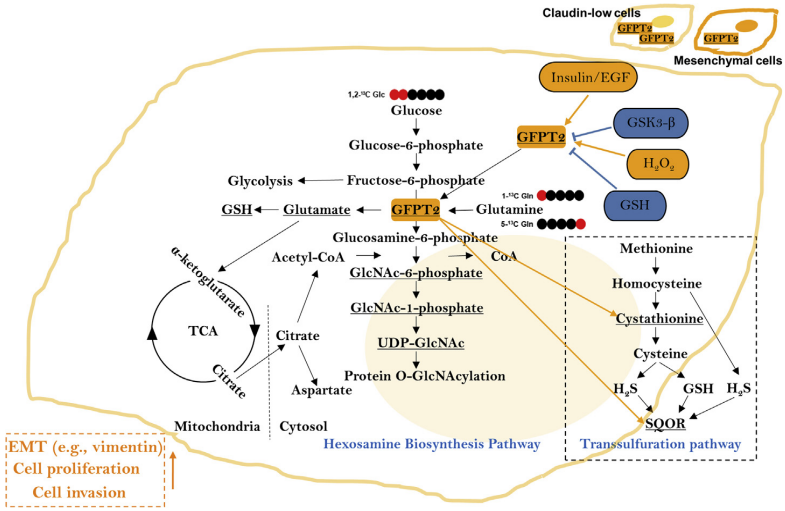


FIG. 9. **Summary of the study.** The hexosamine biosynthesis pathway (HBP) and transsulfuration pathway, two main pathways associated with GFPT2 in this study, were illustrated. HBP is closely associated with the central carbon metabolism. The underscored metabolites were measured in this study. We used 1,2-^{13}C Glc, 1-^{13}C Gln, and 5-^{13}C Gln in the ^{13}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_2O_2 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_2S originates from the transsulfuration pathway, and knockdown of *GFPT2* resulted in decreases to the pathway intermediate, cystathionine, suggesting that the production of H_2S could be hampered. SQOR utilizes H_2S as substrate, and decreased SQOR with *GFPT2* knockdown supports the hampered H_2S production. However, a solid relationship between GFPT2 and H_2S 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_2S and further with mitochondrial functions. H_2S signaling has recently been reported to facilitate EMT (79, 80). However, the exact role of H_2S 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_2S 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_2S level that impacts SQOR and mitochondria homeostasis. The protective effects of GFPT2 from oxidative stress may thus be attributed in part to changes in H_2S-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-β and casein kinase. We interrogated GSK3-β 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-β phosphorylation was higher in D492HER2 suggesting inactivation of GSK3-β *via* phosphorylation (90). GSK3-β RNA and protein levels were increased in D492, indicating more active GSK3-β kinase activity in D492. Knockdown of *GSK3-β* in the D492 cells upregulated GFPT2 expression suggesting that GSK3-β 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 downregulated in epithelial D492 by GSK3-β 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-β 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-β 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 kkpmzoav00 (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 fhlawvfolp (iBAQ).

Supplemental data—This article contains supplemental data (43).

Acknowledgments—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. The graphical abstract was created with BioRender.com. This work was funded by Icelandic Center for Research (RANNÍS, 163254-052) and Göngum 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, Dermicidin; 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-β, Glycogen synthase kinase 3 beta; GSSG, Oxidized glutathione; H_2O_2, Hydrogen peroxide; H_2S, 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-κB, 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; PKCα, 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-κB 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-β, Transforming growth factor beta; TNFα, 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 from, October 20, 2021
Published, MCPRO Papers in Press, December 17, 2021, https://doi.org/10.1016/j.mcpro.2021.100185

REFERENCES

1. Britt, K. L., Cuzick, J., and Phillips, K. A. (2020) Key steps for effective breast cancer prevention. *Nat. Rev. Cancer* **20**, 417–436
2. Weigelt, B., Peterse, J. L., and van 't Veer, L. J. (2005) Breast cancer metastasis: Markers and models. *Nat. Rev. Cancer* **5**, 591–602
3. 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
4. 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
5. Nieto, M. A., Huang, R. Y., Jackson, R. A., and Thiery, J. P. (2016) EMT: 2016. *Cell* **166**, 21–45
6. Lambert, A. W., Pattabiraman, D. R., and Weinberg, R. A. (2017) Emerging biological principles of metastasis. *Cell* **168**, 670–691
7. Rivenzon-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
8. Barnabas, G. D., Lee, J. S., Shami, T., Harel, M., Beck, L., Selitrennik, M., Jerby-Arnon, 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
9. Halldorsson, S., Rohatgi, N., Magnusdottir, M., Choudhary, K. S., Gudjonsson, T., Knutsen, E., Barkovskaya, A., Hilmarsdottir, B., Perander, M., Mælandsmo, 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
10. Lucenay, K. S., Doostan, I., Karakas, C., Bui, T., Ding, Z., Mills, G. B., Hunt, K. K., and Keyomarsi, 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
11. Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., Verfaillie, C., Grünewald, 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
12. Lv, 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
13. Bergers, G., and Fendt, S. M. (2021) The metabolism of cancer cells during metastasis. *Nat. Rev. Cancer* **21**, 162–180
14. Ingthorsson, S., Andersen, K., Hilmarsdottir, B., Maelandsmo, G. M., Magnusson, M. K., and Gudjonsson, T. (2016) HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. *Oncogene* **35**, 4244–4255
15. Sigurdsson, V., Hilmarsdottir, B., Sigmundsdottir, H., Fridriksdottir, A. J., Ringnér, M., Villadsen, R., Borg, A., Agnarsson, B. A., Petersen, O. W., Magnusson, M. K., and Gudjonsson, T. (2011) Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One* **6**, e23833
16. Briem, E., Ingthorsson, 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
17. 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
18. Ackland, M. L., Newgreen, D. F., Fridman, M., Waltham, M. C., Arvanitis, A., Minichiello, J., Price, J. T., and Thompson, E. W. (2003) Epidermal growth factor-induced epithelio-mesenchymal transition in human breast carcinoma cells. *Lab. Invest.* **83**, 435–448
19. Xu, J., Lamouille, S., and Derynck, R. (2009) TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* **19**, 156–172
20. 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
21. 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
22. 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 β-catenin. *Pathol. Res. Pract.* **215**, 152681
23. Karvelsson, S. T., Sigurdsson, A., Seip, K., Grinde, M. T., Wang, Q., Johannsson, F., Mælandsmo, G. M., Moestue, S. A., Rolfsson, O., and Halldorsson, S. (2021) EMT-derived alterations in glutamine metabolism sensitize mesenchymal breast cells to mTOR inhibition. *Mol. Cancer Res.* **19**, 1546–1558
24. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362
25. 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
26. 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
27. Cox, J., Neuhauser, N., Michalski, A., Scheltema, 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
28. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed

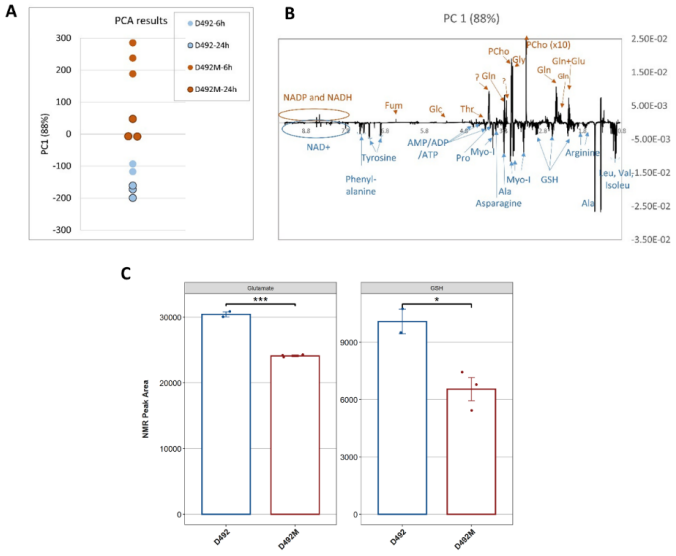
normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526
29. Rolfsson, Ó., Johannsson, F., Magnusdottir, M., Paglia, G., Sigurjonsson, Ó. E., Bordbar, A., Palsson, S., Brynjólfsson, S., Guðmundsson, S., and Palsson, B. (2017) Mannose and fructose metabolism in red blood cells during cold storage in SAGM. *Transfusion* **57**, 2665–2676
30. Gu, Z., Eils, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849
31. Galili, 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 (prote)omics data. *Nat. Methods* **13**, 731–740
33. Ulgen, E., Ozisik, O., and Sezerman, 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., Sidiropoulos, 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., Junge, 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., Wiznerowicz, 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. Koleti, 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 precalculated 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., Bierie, 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 Villén, 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., García 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-κB 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-κB upregulates glutamine-fructose-6-phosphate transaminase 2 to promote migration in non-small cell lung cancer. *Cell Commun. Signal.* **17**, 24
54. 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., *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**, 3445–3457
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., Karlan, 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 epithelialmesenchymal 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., Tonon, 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., 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., *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., Kasiri, S., *et al*. (2020) The hexosamine biosynthesis pathway is a targetable liability in KRAS/LKB1 mutant lung cancer. *Nat. Metab.* **2**, 1401–1412
75. Eiriksson, 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-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J. Clin. Invest.* **114**, 569–581
79. Ascencao, K., Dilek, N., Augsburger, 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-1α 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-β1-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., Zara, 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) Farrerol directly targets GSK-3β to activate Nrf2-ARE pathway and protect EA.hy926 cells against oxidative stress-induced injuries. *Oxid. Med. Cell Longev.* **2020**, 5967434
88. Riis, S., Murray, J. B., and O'Connor, R. (2020) IGF-1 signalling regulates mitochondria dynamics and turnover through a conserved GSK-3β-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-GSK3beta 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. Beurel, 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-3beta 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., Inuganti, 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., Yildirir, G., Henry, R., *et al*. (2019) O-GlcNAcylation is required for mutant KRAS-induced lung tumorigenesis. *J.Clin. Invest* **128**, 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., Guinamard, R., de Tenbossche, C. P., Verdeil, G., Van den Eynde, B., and Schmitt-Verhulst, A. M. (2012) Epithelial-mesenchymal-transition-like and TGFbeta pathways associated with autochthonous inflammatory melanoma development in mice. *PLoS One* **7**, e49419
99. Sitole, B. N., and Mavri-Damelin, D. (2018) Peroxidasin is regulated by the epithelial-mesenchymal transition master transcription factor Snai1. *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 NNMT 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

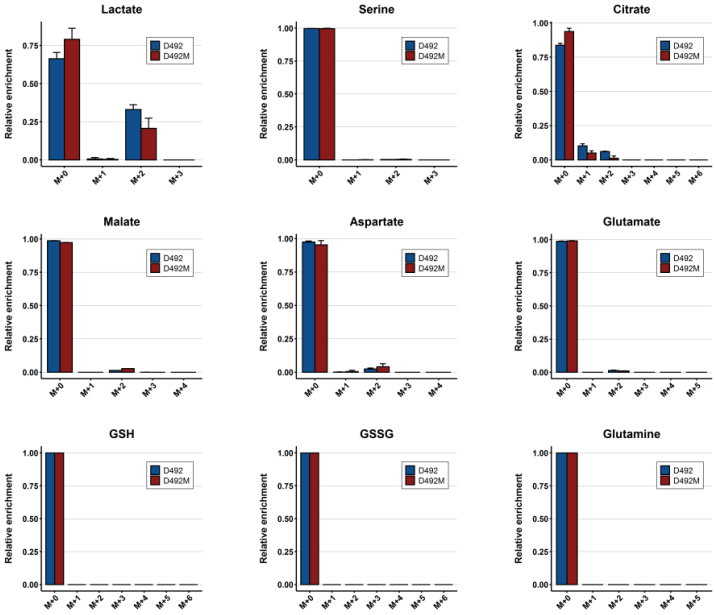
102. Bonet, F., Dueñas, Á., López-Sánchez, C., García-Martínez, V., Aránega, A. E., and Franco, D. (2015) MiR-23b and miR-199a impair epithelial-to-mesenchymal transition during atrioventricular endocardial cushion formation. *Dev. Dyn. off. Publ. Am. Assoc. Anatomists* **244**, 1259–1275
103. Alonso, 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., Feliu, 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

Appendix I

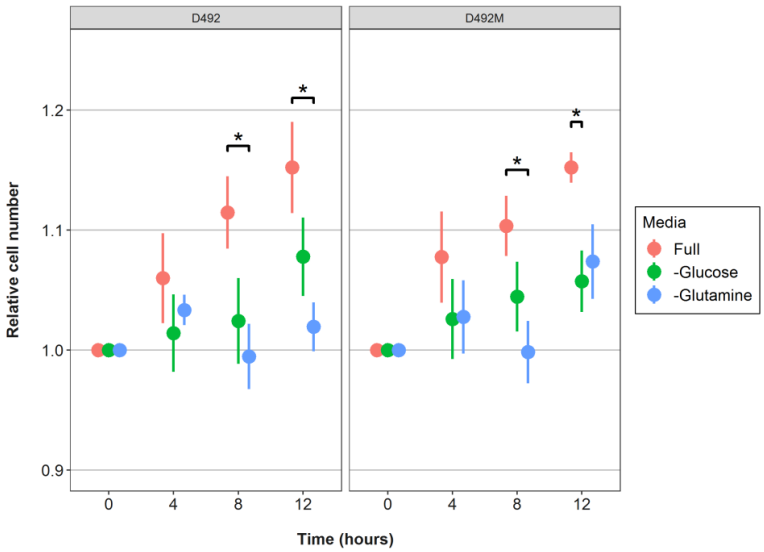
Supplementary files for Paper I



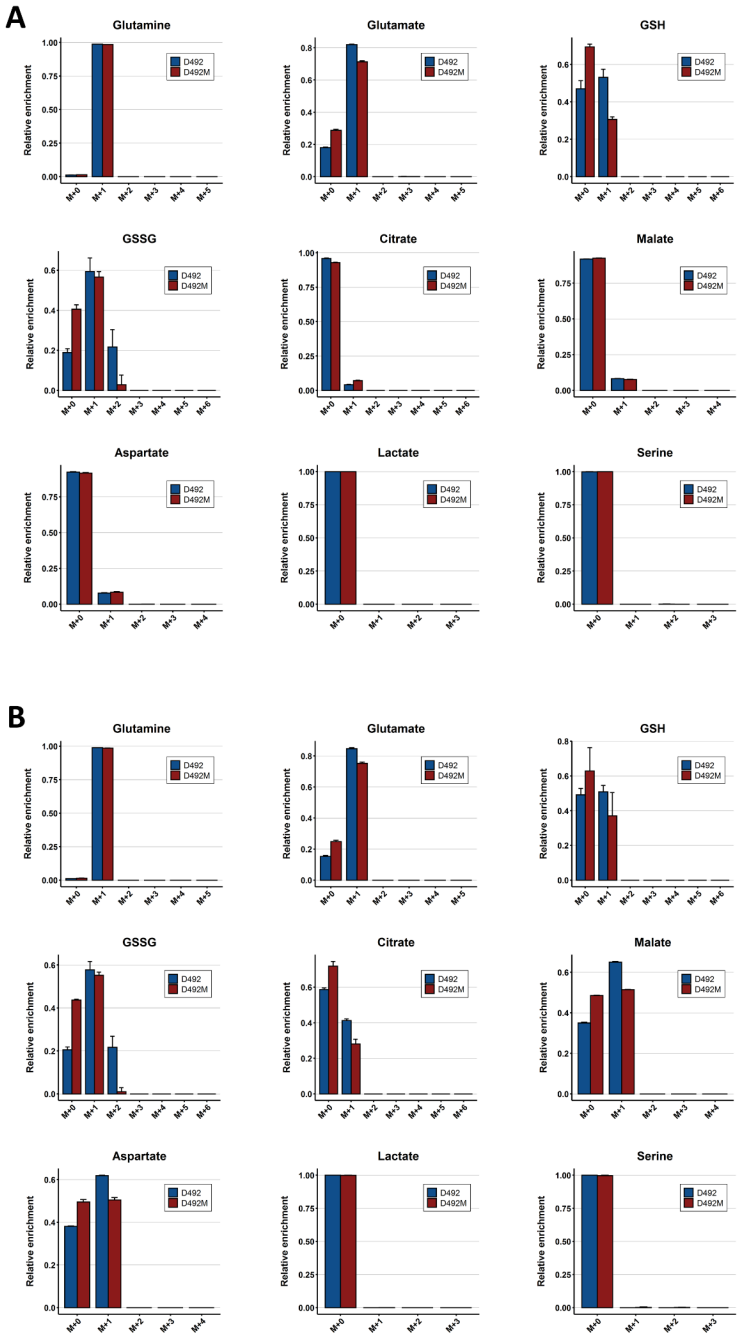
Supplementary figure 1. *NMR data from 1H and 13C labelling experiments of D492 and D492M*
A) PC 1 scores after Principal Component Analysis of 1H NMR data and B) Corresponding loading plot from PC 1[£]*. C) NMR peak area after incubation with 1-13C-glutamine for 6 hours. The metabolites show integrated peak areas of 1-13C glutamate and 1-13C glutathione (GSH).* [£]*: Peak assignment: a: isoleucine, b: leucine, c: valine, d: lactate, e: alanine, f: arginine, g: glutamine, h: glutamate, i: glutathione, j: phosphocholine (10 times higher than shown), k: myo-inositol, l: glycine, m: asparagine, n: proline, o: amp, p: adp, q: atp: r: threonine, s: glucose, t: fumarate, u: tyrosine, v: phenylalanine, w: NAD+, x: NADP, y: NADH.*



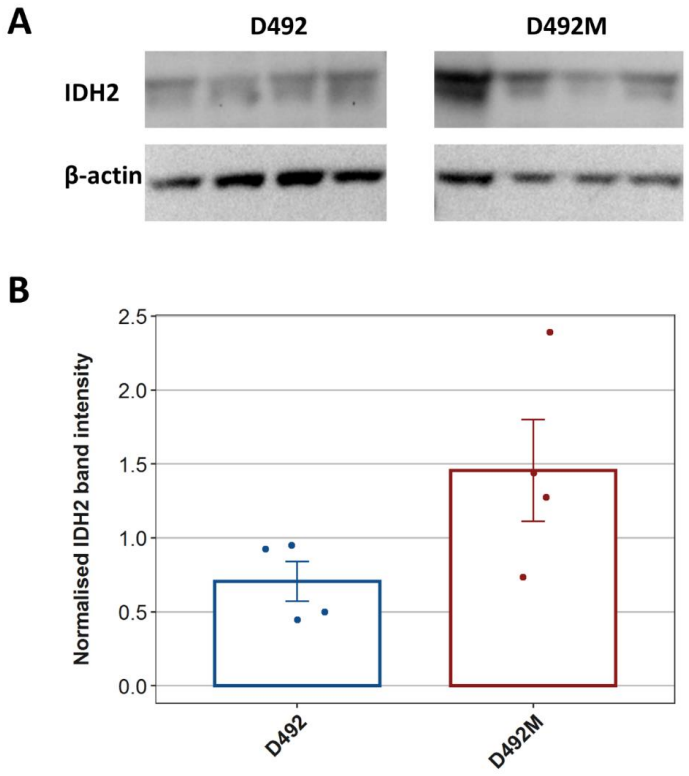
Supplementary figure 2. *UPLC-MS label enrichment in various metabolites after culturing of D492 and D492M with 1,2-13C-glucose for 6 hours. Results are shown as mean + standard deviation (n = 3).*



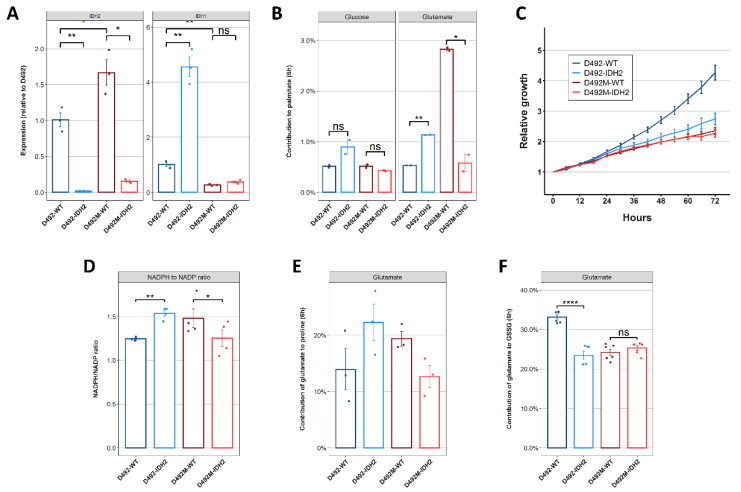
Supplementary figure 3. *Dependency of D492 and D492M on glucose and glutamine. Cell number of D492 and D492M in full H14 medium, H14 without glucose and H14 without glutamine. Values are presented as mean + standard error (n≥6). Asterisks represent significance of pairwise comparison between full medium and the depleted medium.*



Supplementary figure 4. *UPLC-MS label enrichment in various metabolites after culturing of D492 and D492M with A) 1-^{13}C-glutamine and B) 5-^{13}C-glutamine for 6 hours. Results are shown as mean + standard deviation (n = 3).*

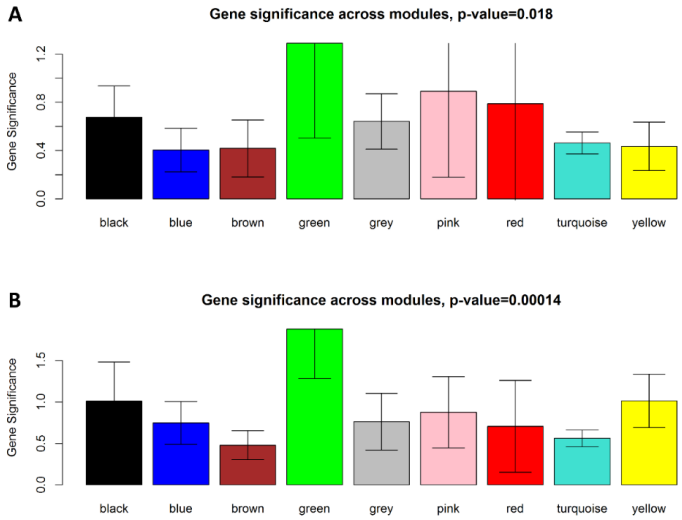


Supplementary figure 5. *Western blot showing levels of IDH2 in D492 and D492M. A) IDH2 and β-actin intensity levels in D492 and D492M. B) IDH2 band intensity normalized to the β-actin loading control in D492 and D492M (n=4). Student's two-tailed t-test was used to estimate significance of results. The difference between D492 and D492M in normalized IDH2 band intensity was not significant (p = 0.08).*



Supplementary figure 6. *IDH2 knockdown effects on D492 and D492M. A) Real-Time PCR from D492, D492M and their IDH2 -silenced cell lines showing the gene expression levels of IDH2 and IDH1. B) Effect of IDH2 silencing on the reductive contribution of glutamate and 1,2-^{13}C-glucose to palmitate in D492 and D492M. C) Proliferation of D492 and D492M WT*

and IDH2-knockdown cell lines (mean + sem, n=8). D) Effect of IDH2 silencing on the NADPH/NADP$^+$ ratio in D492 and D492M. E) Effect of IDH2 silencing on the contribution of 1-^{13}C-glutamine to proline. F) Effect of IDH2 silencing on the contribution of 1-^{13}C-glutamine to oxidized glutathione (GSSG). Student's two-tailed t-test (with Benjamini-Hochberg adjustment for multiple comparisons) was used to estimate significance of results.



Supplementary figure 7. *Summarized significance (-log(p-value)) of drug module members Spearman correlation to levels of A) Reduced glutathione (GSH) and B) Oxidized glutathione (GSSG). The p-values in the title represents significance values from Kruskal-Wallis tests for differences in significance value distributions between modules.*

Supplementary file 1. *An R-script for the weighted correlation network analysis. This file is a .R file which can be opened in RStudio.*

```
# SUPPLEMENTARY FILE FOR KARVELSSON ET AL.
################################################################################

# Load all necessary libraries:
library(WGCNA) # For network analysis
library(rcellminer) # For extraction of drug sensitivity data from NCI60
library(readxl) # To read Excel files
library(tidyr) # For the drug MOA enrichment test
library(dplyr) # For the drug MOA enrichment test
library(siggitRausti) # To sample the drug annotation terms


### LOAD AND PREPROCESS DATASETS ###

# Load data from Ortmayr et al. (https://www.nature.com/articles/s41467-019-09695-9)
# Note: This is supplementary data available from the paper in the link.
met_means <- read_excel('41467_2019_9695_MOESM2_ESM.xlsx','Met_alpha','A1:BG2185')
met_sds <- read_excel('41467_2019_9695_MOESM2_ESM.xlsx','Met_alpha_SE','A1:BG2185')
```

```
colnames(met_means) <- unlist(as.character(met_means[3,]))
colnames(met_sds) <- unlist(as.character(met_sds[3,]))
met_means <- met_means[-c(1:3),]
met_sds <- met_sds[-c(1:3),]

# Change name of tissues so that they will match:
met_means2 <- met_means
drugAct <- exprs(getAct(rcellminerData::drugData))
drugAct <- data.frame(drugAct)

drugAct_2 <- drugAct
colnames(met_means2) <- gsub('Breast_','BR.',colnames(met_means2))
colnames(met_means2) <- gsub('CNS_','CNS.',colnames(met_means2))
colnames(met_means2) <- gsub('Colon_','CO.',colnames(met_means2))
colnames(met_means2) <- gsub('Lung_','LC.',colnames(met_means2))
colnames(met_means2) <- gsub('Melanoma_','ME.',colnames(met_means2))
colnames(met_means2) <- gsub('Ovarian_','OV.',colnames(met_means2))
colnames(met_means2) <- gsub('Prostate_','PR.',colnames(met_means2))
colnames(met_means2) <- gsub('Renal_','RE.',colnames(met_means2))
colnames(met_means2) <- gsub('\\.','',colnames(met_means2))
#rownames(heyhey2) <- gsub('\\.','',rownames(heyhey2))
colnames(drugAct_2) <- gsub('\\.','',colnames(drugAct_2))

# Extract only FDA-approved drugs:
drugAnnot <- getFeatureAnnot(rcellminerData::drugData)[["drug"]]
tmpDA <- drugAnnot[drugAnnot$FDA_STATUS != "-", c("NSC", "FDA_STATUS")]
tmpDrugAct <- drugAct_2[rownames(drugAct_2) %in% tmpDA$NSC,]
tmpDrugAct <- tmpDrugAct[grep('FDA',tmpDA$FDA_STATUS),]

# Create a metabolite dataframe with same order as drug dataframe
metabolite_dataframe <- data.frame(matrix(NA,nrow = 54,ncol = nrow(met_means2)))
rownames(metabolite_dataframe) <- colnames(met_means2)[6:ncol(met_means2)]
for (i in 1:nrow(met_means2)){
  metabolite_dataframe[,i] <- as.numeric(as.character(unlist(met_means2[i,6:ncol(met_means2)])))
  colnames(metabolite_dataframe)[i] <- make.names(met_means2$Annotation[i])
}

# Now add the drug of interest:
metabolite_dataframe$Drug_efficacy <- rep(NA,nrow(metabolite_dataframe))
for (i in 1:nrow(metabolite_dataframe)){
  id_in_glut <- which(colnames(tmpDrugAct) %in% rownames(metabolite_dataframe)[i])
  if (length(id_in_glut) == 0){
    metabolite_dataframe$Drug_efficacy[i] <- NA
  } else {
    metabolite_dataframe$Drug_efficacy[i] <- tmpDrugAct[which(rownames(tmpDrugAct) %in% '758706'),id_in_glut]
  }
```

```
}

cell_type_vec <- rep(0,nrow(metabolite_dataframe))
for (k in 1:nrow(metabolite_dataframe)){
  cell_type_vec[k] <- substr(rownames(metabolite_dataframe)[k],start=1,stop=2)
}
metabolite_dataframe$Cell_type = cell_type_vec

# Rearrange data
metabolite_dataframe <- metabolite_dataframe[,c(2182,2183,1:2181)]
metabolite_dataframe$Cell_type <- as.factor(metabolite_dataframe$Cell_type)



# Now take out drugs with more than 10% missing data:
NetworkData <- data.frame(t(tmpDrugAct))



### CONSTRUCT A CORRELATION NETWORK ###

gsg = goodSamplesGenes(NetworkData, verbose = 3)
if (!gsg$allOK)
{
  # Optionally, print the gene and sample names that were removed:
  if (sum(!gsg$goodGenes)>0)
    printFlush(paste("Removing genes:", paste(names(NetworkData)[!gsg$goodGenes], collapse = ", ")))
  if (sum(!gsg$goodSamples)>0)
    printFlush(paste("Removing samples:", paste(rownames(NetworkData)[!gsg$goodSamples], collapse = ", ")))
  # Remove the offending genes and samples from the data:
  NetworkData = NetworkData[gsg$goodSamples, gsg$goodGenes]
}

# Plot the cell lines according to how they are affected by drugs:
sampleTree = hclust(dist(NetworkData,method = 'euclidean'),method = 'complete')
sizeGrWindow(12,9)
par(cex = 0.6);par(mar = c(0,4,2,0))
plot(sampleTree, main = "Sample clustering of trauma patients", sub="", xlab="", cex.lab = 1.5,cex.axis = 1.5, cex.main = 2)
powers = c(c(1:10), seq(from = 12, to=20, by=2))

# Call the network topology analysis function
sft = pickSoftThreshold(NetworkData, powerVector = powers, verbose = 5,networkType = 'signed')

# Plot the results:
sizeGrWindow(9, 5)
par(mfrow = c(1,2))
cex1 = 0.9
# Scale-free topology fit index as a function of the soft-thresholding power
```

```
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],xlab="Soft Threshold (power)",ylab="Scale Free Topology Model Fit,signed R^2",type="n",main = paste("Scale independence"))
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],labels=powers,cex=cex1,col="red")
abline(h=0.85,col="red")


# Mean connectivity as a function of the soft-thresholding power
plot(sft$fitIndices[,1], sft$fitIndices[,5],xlab="Soft Threshold (power)",ylab="Mean Connectivity", type="n",main = paste("Mean connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5], labels=powers, cex=cex1,col="red")

# Lets have the value at 9:
net = blockwiseModules(NetworkData, power = 9,networkType = 'signed',corType = 'pearson',TOMType = 'signed',
                       deepSplit = 4,maxBlockSize = 1000,minModuleSize = 5,mergeCutHeight = 0.65)

mergedColors = net$colors

# Plot the dendrogram and the module colors underneath (for a single block)
plotDendroAndColors(net$dendrograms[[1]], mergedColors[net$blockGenes[[1]]],"Module colors",
                    dendroLabels = FALSE, hang = 0.03,addGuide = TRUE, guideHang = 0.05)


# Calculate eigenvectors:
MEList = moduleEigengenes(NetworkData, colors = mergedColors)
MEs = MEList$eigengenes
# Remove grey module:
table(net$colors)
table(mergedColors)
MEs <- removeGreyME(MEs,greyMEName = 'MEgrey')

# Cluster and plot module eigenvectors:
MEDiss = 1-cor(MEs);
METree = hclust(as.dist(MEDiss), method = "ward.D2");
# Plot the result
sizeGrWindow(7, 6)
plot(METree, main = "Clustering of module eigenmetabolites",
    xlab = "", sub = "")


### GET THE ENRICHED MECHANISM OF ACTION FOR ALL DRUG MODULES ###

# Get the names and mechanism of action (MOA) for compounds with MOA entries
knownMoaDrugs <- unique(c(getMoaToCompounds(), recursive = TRUE))
knownMoaDrugInfo <- data.frame(NSC=knownMoaDrugs, stringsAsFactors = FALSE)

# Process all NSCs
knownMoaDrugInfo$MOA <- sapply(knownMoaDrugInfo$NSC, getMoaStr)
```

```
# Split on commas to get paired MOAs and NSC annotations:
mydat <- knownMoaDrugInfo %>%
  mutate(MOA2 = strsplit(as.character(MOA), ",")) %>%
  unnest(MOA2)

# Take out PK: drugs, since they mess with the enrichment test and are not mapped to any MOA:
id_shorten <- grep('PK:',mydat$MOA2)
mydat$MOA2[id_shorten] <- substring(mydat$MOA2[id_shorten],4)

# Create a function that takes n drugs, and calculates the permutation p-value to find enriched MOA terms:
enriched_drugs <- function(drug_list,drug_annotations_mat,n_samples){
  # the drug list should be a character vector of drugs with names in first column and annotations in second
  n_drugs <- length(drug_list)
  drug_annotations_mat_new <- drug_annotations_mat[which(drug_annotations_mat[,1] %in%
unique(drug_annotations_mat[,1])),]
  sampled_mat <- sample_vector(drug_annotations_mat_new[,2],n_drugs,n_samples,replicate = T)
  drug_counts <- data.frame(unique(drug_annotations_mat[,2]))
  drug_counts$r_vals <- NA
  colnames(drug_counts)[1] <- 'drugs'
  drug_counts$my_count <- NA
  my_drug_counts <- drug_annotations_mat[which(drug_annotations_mat[,1] %in% drug_list),2]
  for (i in 1:nrow(drug_counts)){
    null_dist <- as.numeric(colSums(sampled_mat == drug_counts$drugs[i]))
    drug_counts$my_count[i] <- length(which(my_drug_counts == drug_counts$drugs[i]))
    r_val <- length(which(null_dist >= drug_counts$my_count[i]))
    drug_counts$r_vals[i] <- r_val
  }
  drug_counts$p_vals <- (drug_counts$r_vals + 1)/(n_samples + 1)
  drug_counts$adj_p_vals <- p.adjust(drug_counts$p_vals,method = 'bonferroni')
  return(drug_counts)
}

# Preprocess the NSC and MOA dataframe:
drug_annotations <- data.frame(mydat[,c(1,3)])
fda_drugs <- substring(colnames(NetworkData),2)
drug_annotations <- drug_annotations[which(drug_annotations$NSC %in% fda_drugs),]
drug_annotations <- unique(drug_annotations)

# Get the list of drugs for all modules and check enriched mechanisms of action:
my_module = 'black' # Test all different module colors
module_drugs = names(NetworkData)[mergedColors==my_module]
drug_list <- substring(module_drugs,2)

# Test the enrichment of MOA:
```

```
enrichment_dataframe <- enriched_drugs(drug_list,drug_annotations,10000)

# Are there any significant terms enriched after adjusting for multiple tests?
Drug_MOA_Key[enrichment_dataframe$drugs[which(enrichment_dataframe$adj_p_vals < 0.05)], ]



### EXAMINE ASSOCIATION OF METABOLITES TO DRUG MODULES ###

# Create the trait matrix with selected metabolite (glutathione):
id_metabolite <- rep(NA,nrow(MEs))
for (i in 1:nrow(MEs)){
  id_temp <- which(rownames(metabolite_dataframe) %in% rownames(MEs)[i])
  if (length(id_temp) != 0){
    id_metabolite[i] <- id_temp
  } else {
    id_metabolite[i] <- NA
  }
}

# Create a trait matrix based on specific metabolites (GSH, GSSG, UDP-glucuronate and S-adenosylmethionine)
id_mets_chosen <- c(589,1317,1257,859)
traits <- data.frame(metabolite_dataframe[id_metabolite,id_mets_chosen])
colnames(traits) <- c('GSH','GSSG','UDP-Glucuronate','SAM')
nSamples = nrow(NetworkData)
moduleTraitCor = WGCNA::cor(MEs, traits, use = "p",method = 'pearson')
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples)

# Label the modules according to MOA enrichment tests:
y_labels_for_image <- c('Tyrosine kinase inh. 1','Tyrosine kinase inh. 2','ALK/CDK inh.','mTOR/STK inh.','Various',
                        'Hormones','DNA-damaging agents','MAPK inh.')



# Plot the correlation:
textMatrix = paste(signif(moduleTraitCor, 2), "\n(",signif(moduleTraitPvalue, 1), ")", sep = "")
dim(textMatrix) = dim(moduleTraitCor)
png(file="Heatmap_drug_module_GSH_JAN2021.png",
    width=1150, height=750,res = 100)
par(mar = c(10, 14, 3, 3))
# Display the correlation values within a heatmap plot
labeledHeatmap(Matrix = moduleTraitCor,xLabels = names(traits),
               yLabels = names(MEs),
               yColorLabels = FALSE,
               colors = blueWhiteRed(10),textMatrix = textMatrix,
               ySymbols = y_labels_for_image,
               cex.lab.x = 1.5,font.lab.x = list('face' = 2),
               font.lab.y = list('face' = 2),
```

```
        setStdMargins = FALSE,cex.text = 1.5,
        zlim = c(-1,1),
        main = paste(""))
dev.off()

# The green module is sign. negatively correlated with glutathione. What drugs are in there?
my_module = 'yellow'
magenta = names(NetworkData)[mergedColors==my_module]
magenta
getDrugName(substring(magenta,2))



### PLOT THE NETWORK USING IGRAPH ###

library(igraph)
# Put the size of the nodes as the correlation to glutathione (GSSG):
net$correlation <- 10^-(sapply(NetworkData, function(x) cor(x,as.numeric(metabolite_dataframe[id_metabolite,1317]),
                                    use = 'pairwise.complete.obs',method = 'pearson')))
net$correlation[which(is.na(net$correlation))] <- mean(net$correlation,na.rm=T)
TOM = TOMsimilarityFromExpr(NetworkData, power = 9,TOMType = 'signed',networkType = 'signed',corType = 'pearson')
adj <- TOM[-which(net$colors == 'grey'),-which(net$colors == 'grey')]
adj[adj > 0.025] <- 1
adj[adj != 1] <- 0
network <- graph.adjacency(adj)
network <- simplify(network)
V(network)$color <- net$colors[-which(net$colors == 'grey')]
V(network)$size <- net$correlation[-which(net$colors == 'grey')]*5
V(network)$label <- ""
E(network)$arrow.mode <- 0
par(mar=c(0,0,0,0))
network <- delete.vertices(network, degree(network)==0)
#png('Graph_paper2_JAN2021.png',width=6.8, height=6.8,units = 'in',res=300)
plot.igraph(network, layout=layout.fruchterman.reingold(network), edge.arrow.size = 0.05)
legend(x=-1.4, y=-0.5, c("DNA-damaging agents","Hormones", "Various","Tyrosine kinase inh. 2",
                "mTOR/STK inh.","ALK/CDK inh.","Tyrosine kinase inh. 1","MAPK inh."), pch=21,
    col="#777777", pt.bg= unique(net$colors)[-3], pt.cex=2, cex=.8, bty="n", ncol=1)
#dev.off()
```

Supplementary files for Paper II

1. Materials and methods

1.1. Construction of cell-type specific epithelial and mesenchymal GSMMs

To model the metabolism of D492 and D492M, intracellular reactions within iBreast2886 breast model was constrained using cell-type specific omics data and extracellular uptake and secretion rates were bounded by measured uptake and secretion rates of the two cell lines (**Supplementary figure 4**). In this study, the accuracy of three different omics data in predicting the metabolic phenotypes of D492 and D492M were compared. These omics data were 1) microarray, 2) proteomic and 3) RNA sequencing data. Using each of these datasets, the breast tissue GSMM iBreast2886 was used to create two GSMMs, EPI and MES, representative for the metabolism of D492 and D492M, respectively. The first step in the analysis of all the datasets was the identification of genes and proteins that are downregulated in D492 and D492M to use for intracellular reaction constraining.

1.1.1. Identification of genes/proteins for cell-type specific GSMM creation

Microarray expression data for D492 and D492M[1] were obtained and analysed as described in Halldorsson *et al*.[2]. Briefly, based on a sensitivity analysis, the genes with log-fold relative difference of > +5 in D492M compared to D492 were used to constrain reactions in EPI and genes with log-fold relative difference of < -5 used to constrain reactions in MES. Using the Gene-Protein-Rules (GPRs) in iBreast2886, the reactions associated with the genes were identified. A total of 544 reactions were constrained for both EPI and MES (9.4% of all iBreast2886 reactions).

Proteomic data for the D492 and D492M cells were acquired from Wang *et al.* (Wang et al., submitted). The dataset contains proteomic measurements of both cell lines in triplicate. The levels of all proteins were $\log_2$-transformed and proteins with more than one measurement missing in both cell lines were omitted. The proteomic levels in D492 and D492M were then compared using Student's t-test. The p values were adjusted for multiple comparisons using Storey's FDR-controlling approach[3]. The UniProt identifiers of the proteins were converted to Entrez idenfiers using the R/Bioconductor package *biomaRt*[4,5] to make them consistent with the gene identifiers in the iBreast2886 reconstruction. To facilitate a fair comparison between the proteomic-constrained and microarray constrained GSMMs, a similar number of constrained reactions was desired. A threshold for FDR-adjusted p-values of < 0.05 resulted in a total of 444 constrained reactions for EPI and MES (7.1% of all iBreast2886 reactions).

RNA sequencing (RNA-seq) data for D492 and D492M were obtained from Halldorsson *et al*.[2]. This dataset contains HUGO identifiers and FDR-adjusted p-values. The HUGO identifiers were converted to Entrez identifiers using *biomaRt*. Again, we sought to constrain a similar amount of reactions in the RNA-seq GSMMs as done in the microarray and proteomic-constrained GSMMs. A threshold of FDR-adjusted p-values of < 0.05 resulted in a total of 513 constrained reactions (8.9% of all iBreast2886 reactions).

The lists of reactions to constrain based on the three different omics data are shown in **Supplementary file 1**. After the reactions to constrain were obtained from the omics data, the next step was to apply these constraints to the generic breast reconstruction iBreast2886.

1.1.2 Creation of cell-type specific GSMM creation

A lower bound was imposed on the biomass function in iBreast2886. The value chosen as the lower bound was half of the maximal biomass production rate obtained from a flux balance analysis (FBA)[6] of the model optimizing biomass production. This means that all of the feasible flux vectors in the solution space of iBreast2886 produce some amount of biomass. Prior to adding further constraints based on the different omics data, random sampling of the solution space[7] of iBreast2886 was performed to estimate the distribution of feasible flux values for each reaction. This was achieved using the gpSampler function in the COBRA Toolbox[8], an Artificial Centering Hit-and-Run (ACHR) algorithm which generates flux vectors that satisfy the constraints and the steady-state assumption of the model. For the iBreast2886, 5800 flux vectors were randomly sampled over 32 hours.

Next, the iBreast2886 model was constrained using the omics data of different origin to generate the EPI and MES GSMMs. The genes or proteins significantly upregulated in D492M were linked to reactions using the iBreast2886 GPRs and used to constrain the flux boundaries in the associated reactions to generate the EPI models. For creation of the MES models, the significantly upregulated genes/proteins in D492 were used. Instead of simulating a complete inhibition of reaction activity, the reactions were constrained to the 30th percentile of the their flux values based on the random sampling results from the iBreast2886 GSMM by adjusting the upper and lower bounds of allowable flux values. The upper and lower bounds for the rest of the reactions were set as the maximum and minimum flux values, respectively, from the random sampling of iBreast2886.

After the intracellular reactions had been constrained to create EPI and MES GSMMs using each dataset of different origin, the extracellular uptake and secretion rates in the models were adjusted according to exometabolomic data measured for both D492 and D492M[2]. These targeted exometabolomic data contain measurements of the media concentrations of 39 metabolites over 48 hours of cell culture of both D492 and D492M. The time-dependent concentration values, along with the growth rate of the cells and their dry weight, were used to calculate the cell type-specific uptake and secretion rates of the metabolites as shown in the following equation:

$$v_k = \frac{V\left([M_k]_f - [M_k]_i\right)}{A} \quad (1)$$

Where v_k is the exchange rate for metabolite k, $[M_k]_{48}$ and $[M_k]_0$ are the concentrations of metabolite *k* in the culture media after 48 and 0 hours, respectively, and *A* is the area under the growth curve.

At this stage, we have EPI and MES GSMMs constrained with three different types of omics data and exometabolomic measurements. These are the microarray GSMMs, proteomic GSMMs and RNA-seq GSMMs. As a control, we added another set of EPI and MES GSMMs where no intracellular reactions were constrained with any omics data, but the uptake and secretion rates were constrained on a cell type-specific level using the exometabolomic data. These are referred to as media GSMMs.

A stoichiometric matrix, *S*, contains information about the connection of metabolites and reactions in GSMMs. All EPI and MES models share the same stoichiometry *S* but have different constraints on reaction bounds defined by different abundances of metabolic genes/proteins and different uptake/secretion rates defined by the exometabolome. The workflow of the model construction is outlined in **Supplementary figure 4.**

1.2 Constraint-based modeling and analysis of context-specific GSMMs

If the bounds of reactions in a GSMM are too stringent, then there may not be any feasible flux vectors, i.e. flux vectors that satisfy the constraints of the model. If the cell type-specific EPI or MES models were infeasible once the intracellular and extracellular constraints were applied, the bounds of the reactions were minimally adjusted so that feasibility was achieved. This was achieved using an algorithm previously described[2] which is formulated as the following minimization problem:

$$\text{minimize} \sum_{j \in Rr} (p_j + n_j) \quad (2)$$

$$\mathbf{Sv} = \mathbf{0} \quad (3)$$

$$l_j - n_j \leq v_j \leq u_j + p_j \qquad j \in R_r \quad (4)$$

The objective is to minimize the total adjustment of model reactions (2) by adjusting the relaxation of their upper (p_j) and lower (n_j) bounds while satisyfing the steady state assumption (3). The set of reactions to be relaxed is denoted by R_r (4). The algorithm was implemented in MATLAB using the CVX software for convex programming[9].

Random sampling of the EPI and MES models was performed as described in the previous section for iBreast2886. This was done after placing a lower bound on the biomass function for each model corresponding to 50% of their maximum growth rate based on flux balance analysis (FBA) where the biomass function was optimised. The median of the randomly sampled flux values for each reaction was identified for all cell type-specific GSMMs. In this manner it is possible to generate a single representative flux vector for the EPI and MES models which does not reflect the maximisation of biomass production (like what would be achieved through FBA optimizing biomass production), an unrealistic objective of healthy human cells.

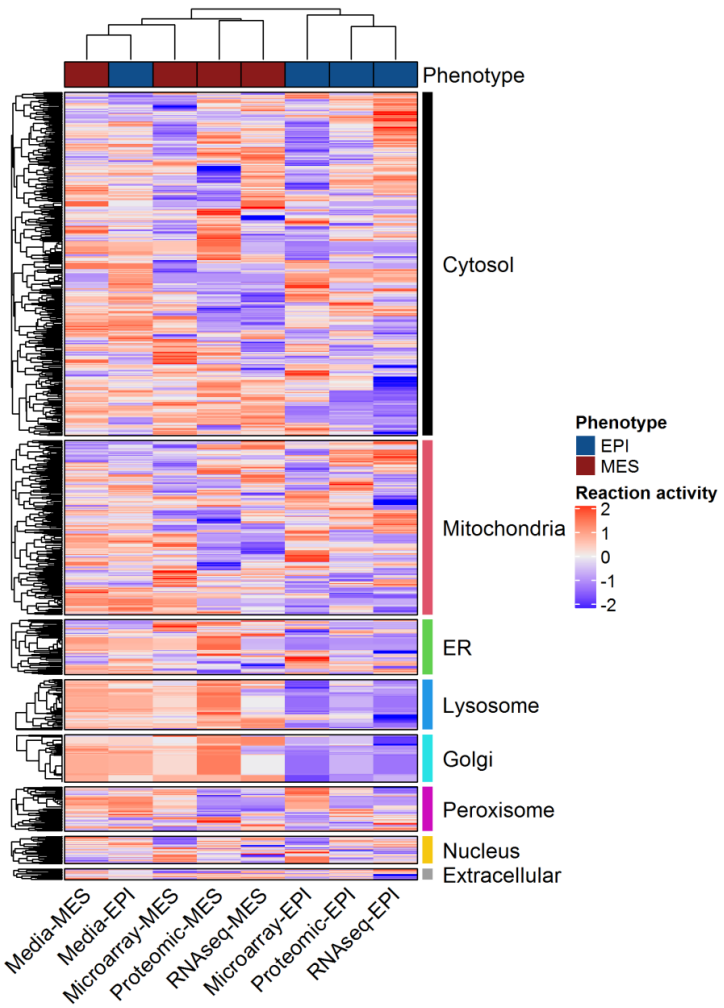
To identify the minimal adjustments needed to divert an EPI flux phenotype towards a MES flux phenotype, representative of the EMT-linked metabolic alterations that occur, a minimization of metabolic adjustment (MOMA)-based algorithm was used. This algorithm finds the minimal alterations needed for an EPI flux vector to become more like a MES flux vector. The optimization problem is formalised in the following manner:

$$minimize \ \|v - v_{MES}\|_1$$

Where v_{EPI} is the median flux vector of EPI from the random sampling (as described above), v_{MES} is the median flux vector of MES and $\|.\|_1$represents the Manhattan norm (1-norm) of a vector. The algorithm returns a list of reactions in EPI whose bounds need relaxation in order to obtain a flux phenotype that resembles that of MES. In our study, we used this list of reactions to perform 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. This was achieved using COBRA Toolbox's fluxEnrichmentAnalysis (FEA) function.

Gene essentiality analysis for all GSMMs was performed using the singleGeneDeletion function within the COBRA Toolbox. This function identifies genes whose inactivation directly affects the predicted growth rate of the models. The genes whose *in silico* deletion caused a complete inibition of biomass production were labeled as essential.

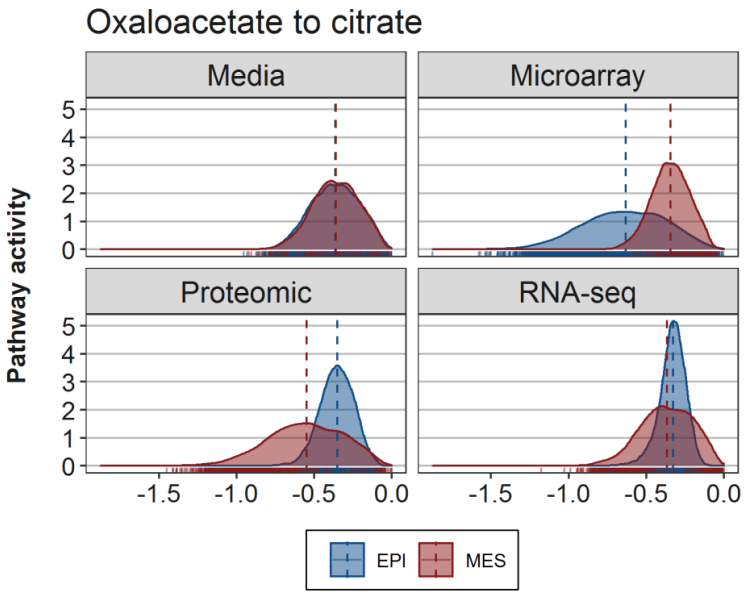
2. Results



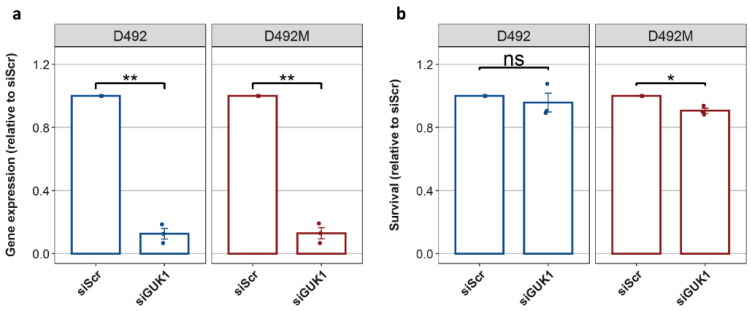
***Supplementary figure 1**. **Heatmap of flux profiles of all GSMMs constrained with different data types.** Hierarchical clustering reveals higher similarity between the EPI and MES phenotypes of differently constrained GSMMs then EPI and MES GSMMs constrained with the same data type. The flux profiles are the median values for each model reaction after random sampling of the GSMMs solution space. Reaction activity is scaled and reactions are shown as members of different subcellular locations (row annotations). The represent the different GSMMs, where blue are EPI GSMMs and red are MES GSMMs The flux vectors are clustered using hierarchical clustering where complete linkage was used as a distance measure.*

Supplementary table 1. Spearman correlation coefficient for relative differences in EPI and MES fluxes between proteomic- and RNA-seq-constrained GSMMs. *The log-fold relative differences in reaction activity between EPI and MES was used as a measure for each reaction in both omics-constrained GSMM pairs. The correlation coefficients (and p-values) between the two omics types were calculated from these measures. The adjusted p-value represents the Bonferroni-adjusted Spearman correlation p-values.*

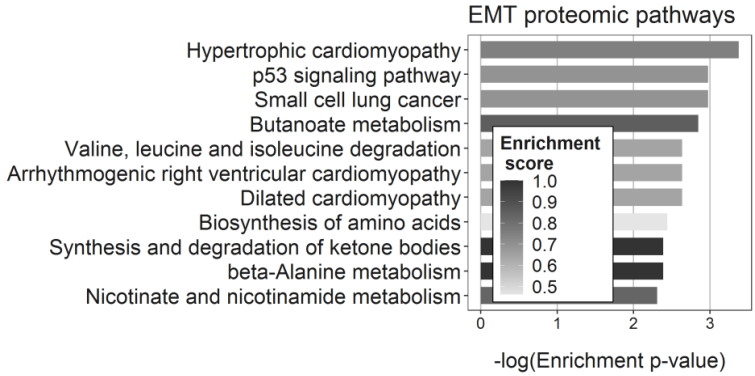
Compartment	Spearman correlation	p-value	Adjusted p-value
Cytosol	0.39	0	0
Mitochondria	0.37	2.2E-12	1.7E-11
Endoplasmic reticulum	0.54	1.1E-09	8.6E-09
Lysosome	-0.1	0.31	1
Golgi apparatus	-0.29	0.0047	0.038
Peroxisome	-0.021	0.85	1
Nucleus	0.29	0.029	0.23
Extracellular	-0.0051	0.98	1



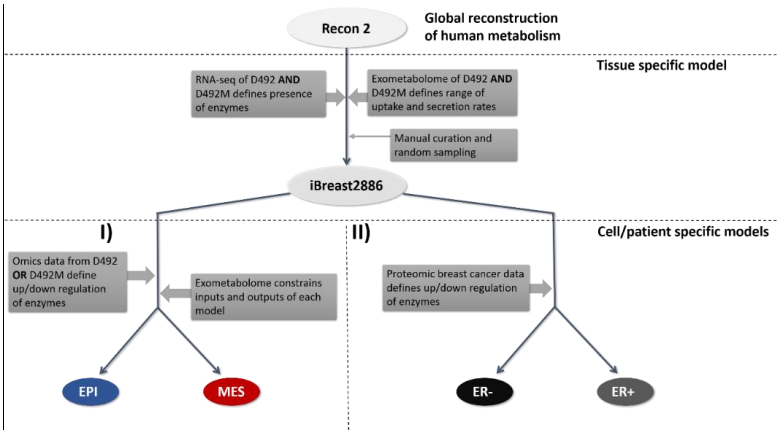
Supplementary figure 2. Density plots of the calculated metabolic route activitiy (MRA) of the condensation of oxaloacetate and acetyl-CoA to generate citrate*. Results are from the total random sampling matrix (n = 5800 flux vectors) for all GSMMs. The blue distributions represent the MRA within the 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 (p < 0.05) based on a Kolmogorov-Smirnov test.*



Supplementary figure 3. Knockdown of GUK1 but not ASS1 inhibits survival of D492M. *A) Expression levels of GUK1 and ASS1 in D492 and D492M after 96 hours of siRNA-mediated knockdown of the genes.) C) siRNA-mediated knockdown of GUK1 and ASS1 and their effects on the 96 hour survival of D492 and D492M. Results in A and B 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.*

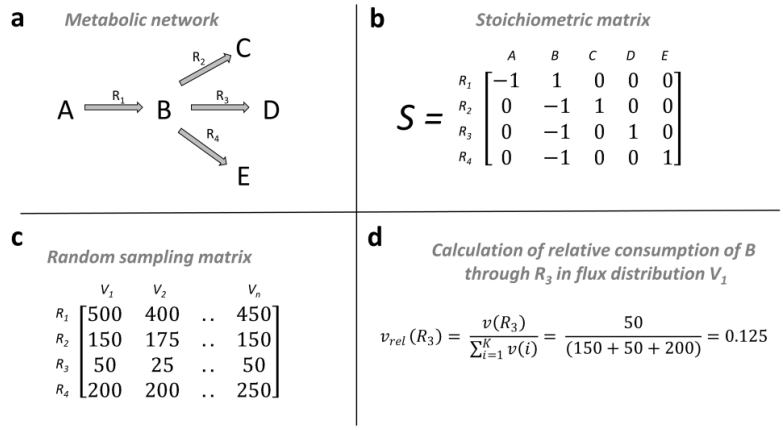


Supplementary figure 4. Enriched KEGG pathways within the differently expressed proteins in D492 and D492M. *After identifying the proteins with FDR-adjusted p-value < 0.01 (n = 588), a KEGG enrichment analysis was performed using a one-sided hypergeometric test. This was performed using the R-package limma, where the background used was the total set of proteins with at least 2 replicates present in either cell line on average (N = 2307). The pathways shown are the ones with p < 0.01. The scale (Enrichment score) represents the fraction of pathway identifiers within the significantly different proteins.*



Supplementary figure 5. Workflow for generating cell- and patient- specific genome-scale metabolic networks. The *networks are descriptive of the metabolism of I) D492 and D492M and II) Breast cancer data from the Tang et al.*[10]*. The generation and the manual curation of iBreast2886 was previously done by Halldorsson et al.*[2]*.*

.



Supplementary figure 6. Explanation of the metabolic route activity calculations of the GSMMs. *a) A simple, metabolic network comprised of metabolites A-E and reactions R_1-R_4. b) A stoichiometric matrix for the metabolic network in A, showing reactions in rows and metabolites in columns. The coefficients represent the abundance of metabolites taking part in each reaction and whether they are consumed or produced. In this example, there are three reactions in which metabolite B is consumed, i.e. in reactions R_2, R_3 and R_4 as these have a -1 for this metabolite. c) An example of a random sampling matrix of the metabolic network in a), subject to the steady state assumption and some additional constraints of the model's reaction (not shown here). Briefly, there are n flux distributions in the sampling matrix (each column). d) Calculation of the relative consumption of B through the R_3 reaction in flux distribution V_1 in the random sampling matrix in c). In this case, the relative contribution of R_3 to the consumption of B is 0.125.*

References

1. Sigurdsson, V. *et al.* Endothelial Induced EMT in Breast Epithelial Cells with Stem Cell Properties. *PLoS One* **6**, e23833 (2011).

2. Halldorsson, S. *et al.* Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett* **396**, 117–129 (2017).

3. D. Storey, J. *A Direct Approach to False Discovery Rates*. vol. 64 (2002).

4. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

5. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

6. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* vol. 28 245–248 (2010).

7. Schellenberger, J. & Palsson, B. O. Use of randomized sampling for analysis of metabolic networks. *J Biol Chem* **284**, 5457–5461 (2009).

8. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).

9. CVX Research, I. {CVX}: Matlab Software for Disciplined Convex Programming, version 2.0. (2012).

10. 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).