



**Identification of novel progression-related
candidate genes in breast cancer**

Arsalan Amirfallah

Thesis for the degree of Philosophiae Doctor

Supervisor/s:

Inga Reynisdóttir

Helga Margrét Ögmundsdóttir

Doctoral committee:

Rosa Bjork Barkardottir, Jill Bargonetti, Óskar Þór Jóhannsson

June 2021



UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES

FACULTY OF MEDICINE

**Skilgreining nýrra brjóstakrabbameinsgena sem
styðja við framvindu æxlismyndunar**

Arsalan Amirfallah

Ritgerð til doktorsgráðu

Leiðbeinandi:

Inga Reynisdóttir

Umsjónarkennari:

Helga Margrét Ögmundsdóttir

Doktorsnefnd:

Rósa Björk Barkardóttir, Jill Bargonetti, Óskar Þór Jóhannsson

Júní 2021



UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES

FACULTY OF MEDICINE

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

© Arsalan Amirfallah 2021

ISBN 978-9935-9586-6-2

Printing by Háskólaprent.

Reykjavik, Iceland 2021

Ágrip

Brjóstakrabbamein er eitt algengasta krabbamein kvenna á heimsvísu. Myndun brjóstaæxla er af margvíslegum toga og m.a. geta breytingar í erfðaefni s.s. stökkbreytingar, genamögnun, litningayfirfærsla, viðsnúningar, innskot og úrfellingar ýtt undir myndun þess. Þrátt fyrir framfarir í einstaklingsmiðaðri meðferð sem byggir á erfðabakgrunni sjúklingsins þá getur meinið komið aftur, jafnvel sex til tíu árum eftir greiningu. Vegna margbreytilegs uppruna brjóstaæxla og mismunandi sameindafræðilegra undirhópa getur verið snúið að finna lífmörk (e. marker) sem spá fyrir um horfur sjúklingsanna. Skilgreining nýrra gena sem taka þátt í æxlisþroska gefur skýrari mynd af því hvernig æxlin taka breytingum, sérstaklega á frumu- og sameindafræðigrunni. Slíka þekkingu mætti nota við að spá fyrir um horfur sjúklingsa, genin mætti nota sem ný lyfjamörk og gæti hún leitt til betri eftirfylgni fyrir sjúklingsinn.

Stökk og snúningar litningabúta innan litnings eða á milli litninga geta leitt til myndunar samrunagena, sem sum hver fá við það illkynja eiginleika. Þau geta orðið ofvirk, e.k. æxlisgen, eða vanvirk, e.k. æxlisbæligen, og einnig myndað ný prótein sem geta haft áhrif á boðleiðir frumunnar. Samrunagen geta myndað samrunaprótein, þ.e. samsett úr hlutum tveggja gena, og geta þau ruglað tjáningu frá genum og microRNA (MIR), sem eru stundum staðsett innan þeirra. Við settum fram þá tilgátu að stakt gen sem kæmi endurtekið fyrir í samrunageni í brjóstaæxlum gæti mögulega verið áhrifagen í brjóstakrabbameinsþróun.

Í verkefninu er stuðst við nýja nálgun til að skilgreina áður óþekkt áhrifagen brjóstakrabbameins. Hún er sú að bera saman samrunagen í brjóstaæxlum og í brjóstakrabbameinsfrumulinum og velja þau sem finnast í báðum. Til að verða fyrir valinu þurftu samrunagenin að uppfylla eftirfarandi skilyrði: 1) vera samsett á sem líkastan hátt í æxlum og frumulinum, 2) vera síendurtekin í æxlum, 3) ekki vera staðsett innan mögnunarsvæðis þekktis æxlisgens nema það væri hluti af samrunageninu, og 4) að genin hefðu virkni sem styður við æxlisþróun. Því næst þurftu genin, sem mynda völdu samrunagenin, að uppfylla tvö skilyrði: 1) sýna háa jákvæða fylgni á milli eintaka gens og mRNA magn þess og 2) sýna fylgni tjáningar gensins við klíníska og meinafræði þætti í sjúklingshópi sem er aðgengilegur í opnum

gagnabanka. Í kjölfarið var Vacuole membrane protein (VMP1) valið til frekari rannsókna. Magn VMP1 mRNA var mælt í brjóstaæxlum tveggja íslenskra hópa. Marktæk tengsl sáust á milli hárrar VMP1 mRNA tjáningar við klíníska og meinafræðipætti, sem tengjast verri horfum, og við skemmri sjúkdómsfría lifun. Tveir erlendir hópar brjóstakrabbameinssjúklinga voru notaðir til að staðfesta niðurstöðurnar og leiddu frekari rannsóknir í ljós að VMP1 tengdist helst skemmri lifun hjá sjúklingum með HER2 jákvæð æxli.

Rannsóknir á VMP1 leiddu til rannsókna á hinu áhrifageni verkefnisins, sem er MIR21. Bæði genin eru staðsett á litningasvæði 17q23.1 og skarast 5' endi MIR21 við 3' enda VMP1. Vegna þessa var hugsanlegt að þau hefðu áhrif á tjáningu hvors annars þó hvort genið um sig hafi eigið stýrisvæði. Þrátt fyrir að hsa-miR-21-5p sé vel þekkt áhrifagen í brjóstakrabbameini (æxlismir) þá er „systkini“ þess hsa-miR-21-3p lítið rannsakað. Við könnuðum áhrif hsa-miR-21-3p í sömu sjúklingahópum og VMP1 og niðurstöður okkar benda til þess að hsa-miR-21-3p er einnig hugsanlegt áhrifagen í þróun brjóstæxla.

Niðurstöður þessara rannsókna sýna að aðferð okkar við að skima samrunagen til að finna hugsanleg áhrifagen í framvindu brjóstæxla virkar. Rannsóknir á virkni VMP1 á sameindafræðigrunni eru hafnar í brjóstakrabbameinsfrumulínum.

Lykilorð: Framvinda brjóstæxla, skimun samrunagena, VMP1, hsa-miR-21-3p

Abstract

Identification of novel progression-related candidate genes in breast cancer

Breast cancer is the most common type of cancer in women worldwide. It is also a highly heterogeneous disease that is characterized by an array of genetic rearrangements, including copy number alterations (CNA), translocations, inversions, insertions, and deletions. Despite progress in targeted therapy, tailored to the patient genetic background, disease relapse is not uncommon, even 6 to 10 years after the initial diagnosis. Because of variety of molecular subtypes and heterogeneity of breast cancer, identifying prognostic and predictive markers has been challenging. Thus, novel approaches to identifying breast cancer genes must be developed.

Inter-chromosomal and intrachromosomal rearrangements can generate fusion genes with oncogenic properties (e.g., oncogene activation, tumor suppressor deletion/downregulation, or chimeric proteins capable of altering cellular pathways); and genes fusions can dysregulate expression of host genes and intragenic miRNAs. We speculated that genes frequently involved in gene fusions in breast tumors are likely associated with breast cancer development and progression.

We developed a novel approach to identify new breast cancer genes. Here, we screened publicly available databanks for fusion genes in breast cancer cell lines and tumors, and candidates had to pass three criteria: 1) the breakpoint must be similar in breast tumors and cell lines, 2) the lesion must be recurrent in tumors, 3) but not located within an amplicon carrying a known oncogene (unless it is part of the fusion), and 4) possess a function supportive of tumorigenesis. Next, the genes that make up the selected fusion gene had to meet two conditions: 1) Show a high positive correlation between gene copy number variations and their mRNA levels, and 2) show a correlation between the expression of the genes and the clinical and pathological aspects of a breast cancer cohort accessible in an open database. Subsequently, Vacuole membrane protein (VMP1) was selected for further research.

By screening two Icelandic breast cancer cohorts and confirming results in two, large, publicly available, breast-cancer cohorts, we identified vacuole membrane protein 1 (VMP1) as a candidate gene involved in the development of breast tumors, particularly the HER2-positive subtype. The role of VMP1 was explored further in HER2 positive BC cell lines and although it did not affect proliferation further studies will reveal whether it affects cellular migration and invasion as well as drug resistance due to its role in autophagy.

The study on VMP1 lead us to the study of a second gene, MIR21. Since there is considerable sequence overlap between the two genes, the possibility remained that expression from the two genes could affect one another. Also, while hsa-miR-21-5p is a well-known oncomir in breast cancer, its “sibling” hsa-miR-21-3p is hardly studied at all. Notably, we found hsa-miR-21-3p (which is transcribed from its own promoter, within intron 10 of VMP1) is a potential marker for breast tumors, confirming the validity of our approach. Data from these two studies showed that screening for fusion genes is a viable method for identifying novel cancer-associated genes. Further, functional cell-based experiments are expected to shed light on the biology of VMP1 and hsa-miR-21-3p, in health and disease.

Keywords: Breast cancer progression, fusion genes screening, VMP1, hsa-miR-21-3p

Acknowledgements

The majority of the work presented in this thesis was carried out at Landspítali, The National University Hospital of Iceland, department of pathology / cell biology unit.

First, I am extremely grateful to my supervisor, Inga Reynisdóttir for her invaluable advice, kindness, continuous support, and patience during my PhD study. Her immense knowledge and plentiful experience have encouraged me in all the years of my PhD in her laboratory. Without her guidance and constant feedback this PhD would not have been achievable.

I also want to express my gratitude to doctoral committee members, Rósa Björk Barkardóttir, Helga Margrét Ögmundsdóttir, Dr. Jill Bargonetti and Óskar Þór Jóhannsson for their input, support, and advice during the years of the project.

I would like to thank both current and previous laboratory members at the department of pathology / cell biology and molecular pathology units: Harpa Lind Björnsdóttir, Eydís Þórunn Guðmundsdóttir, Aðalgeir Arason, Bylgja Hilmarsdóttir, Edda Sigríður Freysteinsdóttir, Katrín Halldórsdóttir and Guðrún Jóhannesdóttir for all their help, support, and friendship over the years of my PhD. I would also like to thank the staff at the department of pathology, especially Sigrún Kristjánsdóttir for preparing paraffin tissue specimens, Bjarni A. Agnarsson for his collaboration in my research project and Jón Gunnlaugur Jónasson, head of the department.

I also want to thank Biomedical Center, Faculty of Medicine and University of Iceland.

I am grateful to Hildur Knútsdóttir for her collaboration and performing of bioinformatic analysis for submitted manuscript.

I appreciate Stefán Sigurðsson and Elísabet Alexandra Frick for sharing the breast cancer cell lines RNAs from their laboratory archive with me.

I also want to thank Diana Colgan for English editing of the thesis.

I want to thank my family in Iran, my friends for all their supports, motivations, and encouragements.

Finally, I thank with love♥ to Mandana and Arsam, my wife and son for their endless love, support, and patience throughout the years of my PhD.

This work was supported by Grants from The Icelandic Centre for Research Fund (152530-051, www.rannis.is), Scientific Fund of Landspítali – The National University Hospital in Iceland (A-2015-039, A-2018-034, www.landspitali.is), grants from Gongum saman (2013, 2017, 2018 <http://www.gongumsaman.is/>) and Icelandic Cancer Society research funds for years 2019 and 2020.

Contents

Ágrip	iii
Abstract	v
Acknowledgements	vii
Contents	ix
List of abbreviations	xiii
List of figures	xv
List of tables	xvii
List of original papers	xviii
Declaration of contribution	xix
1 Introduction	1
1.1 Epidemiology.....	1
1.2 Etiology	1
1.3 Clinical pathology	2
1.3.1 Histograde	2
1.3.2 Staging (TNM)	2
1.3.3 Receptors	2
1.3.4 Molecular subtypes.....	3
1.3.5 Breast cancer diagnosis	4
1.3.6 Therapy.....	4
1.4 Tumor biology	5
1.4.1 Development.....	5
1.4.2 Altered signaling pathways in breast cancer development	5
1.4.3 Known gene mutations in breast cancer	7
1.5 Chromosomal rearrangements in breast cancer	9
1.5.1 Copy number alterations (CNA)	9
1.5.2 Fusion genes	10
1.5.3 Known fusion genes in breast cancer	11
1.6 miRNAs	12
1.6.1 History.....	12
1.6.2 Biogenesis	12
1.6.3 Role of miRNAs in cancer	13
1.6.4 Known miRNAs in breast cancer.....	13
1.7 Breast cancer recurrence:.....	17
1.7.1 Pathophysiology of metastasis	18

1.7.2	Role of drug resistance in breast cancer recurrence.....	19
1.8	Role of autophagy in breast cancer	20
1.9	Breast cancer related data resources	21
2	Aims	23
2.1	Specific aims of the study	23
3	Materials and methods	25
3.1	In silico analysis	25
3.2	Cell lines.....	25
3.3	Cell culture conditions	26
3.4	Breast tumors cohorts	26
3.5	Clinical information and tumor characteristics	27
3.6	DNA/RNA/miRNA extraction from tumors.....	28
3.7	DNA/RNA/miRNA extraction from cell lines	29
3.8	cDNA generation.....	30
3.9	RT-PCR.....	30
3.10	Gel electrophoresis	32
3.11	Sanger sequencing	32
3.12	Gene copy number variation analysis.....	33
3.13	Quantitative reverse-transcription PCR (RT-qPCR)	35
3.14	cDNA generation for miRNA	36
3.15	miRNA quantification with q-PCR	36
3.16	Preparing Whole-Cell Lysates for Immunoblotting:	38
3.17	Western Blot.....	38
3.18	Transfection with siRNA.....	39
3.19	IncuCyte® Live Cell Analysis	39
3.20	Apoptosis assay	40
3.21	Statistical analysis.....	40
4	Results.....	43
4.1	Identification of potential breast cancer genes (Paper 1)	43
4.1.1	Common fusion genes within breast cancer cell lines and tumors.....	43
4.1.2	Verification fusion genes in common between breast tumors and breast cancer cells.....	44
4.1.3	Ten genes identified as putative breast cancer genes	45
4.1.4	Junction site of fusion genes verified.....	46
4.1.5	Role of identified fusion-gene partners in breast cancer development	47
4.1.6	Vacuole membrane protein 1 chosen as a potential breast cancer gene for further studies.....	48

4.2	High expression of the vacuole membrane protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer (Paper 1)	49
4.2.1	VMP1 mRNA levels were high in breast tumors	50
4.2.2	VMP1 mRNA levels were higher in VMP1 amplified tumors.....	51
4.2.3	Co-amplification of VMP1 and ERBB2 genes	52
4.2.4	VMP1 mRNA was higher in HER2 positive tumors.....	53
4.2.5	High VMP1 mRNA expression is associated with poor survival.....	55
4.2.6	VMP1 mRNA correlates with known drivers of the Chromosome 17q23 locus.....	58
4.2.7	The effect of elevated expression of VMP1 mRNA on survival is independent of RPS6KB1, PPM1D, miR21 gene expression	59
4.2.8	HER2 positive patients with high VMP1 had shorter survival.....	60
4.3	Hsa-miR-21-3p is a marker of poor survival in breast cancer patients (manuscript).....	61
4.3.1	Hsa-miR-21-3p was higher in breast tumors.....	62
4.3.2	Hsa-miR-21-3p level is highest in MIR21-amplified tumors.....	62
4.3.3	Hsa- miR-21-3p association with clinic pathologic features.....	64
4.3.4	Hsa-miR-21-3p levels were higher in HER2-positive tumors than HER2-negative tumors	66
4.3.5	Hsa-miR-21-3p associated with histological grade of breast tumors.....	67
4.3.6	High expression of hsa-miR-21-3p as a marker of short survival.....	68
4.3.7	Hsa-miR-21-3p affected survival independently of other clinically pathological features and neighboring genes	71
4.3.8	Hsa-miR-21-3p levels within breast cancer cell lines	72
4.4	Study of VMP1 function in cell lines (unpublished data).....	74
4.4.1	Characteristics of cell lines	74
4.4.2	VMP1 expression was high in ER ⁺ /HER2 ⁺ cell lines	76
4.4.3	VMP1 protein was higher in ER ⁺ /HER2 ⁺ cell lines	78
4.4.4	Silencing of VMP1 and ERBB2 were optimized in BT474 and MDA-MB-361 cells	80

4.4.5	Effect of VMP1 knockdown on cell proliferation and survival.....	81
4.4.6	Effect of VMP1 knockdown on proteins associated with cell adhesion in BT-474 cells.....	84
5	Technical hurdles and considerations.....	87
5.1	Confirmation of fusion genes	87
5.2	Immunohistochemistry	87
5.3	Slow growing cell lines	88
5.4	Transfection of siRNAs	88
5.5	Alternative method for finding link between VMP1 and different proteins.....	89
6	Discussion	91
6.1	Identification of breast cancer genes	91
6.2	Vacuole membrane protein 1 (VMP1).....	93
6.2.1	Biology	93
6.2.2	VMP1's role in breast cancer.....	94
6.2.3	VMP1's role in other cancers.....	95
6.2.4	VMP1 and HER2	96
6.2.5	Discrepant role of VMP1 within cohorts.....	97
6.3	Hsa-miR-21-3p.....	98
6.3.1	The role of hsa-miR-21-3p in breast cancer and other cancer types	99
6.3.2	The discrepant role of hsa-miR-21-3p within cohorts.....	101
6.3.3	Correlation of genes located at chromosome 17q23 locus	102
6.4	VMP1's function in HER2 positive cell lines.....	103
6.5	Implications for future work	105
7	Conclusions.....	107
	References	109
	Paper I.....	135
	Paper II.....	173
	Appendix	213

List of abbreviations

BC	Breast Cancer
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
AJCC	American Joint Committee on Cancer
ER	Estrogen Receptor
PR	Progesterone Receptor
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hormone Receptor
PAM50	Prediction Analysis of Microarray 50
TCGA	The Cancer Genome Atlas
RT-PCR	Reverse Transcription–Polymerase Chain Reaction
UCS	Uterine Carcinosarcoma
LUAD	Lung Adenocarcinoma
OV	Ovarian Tumors
BLCA	Bladder Cancer
CECSC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma
HNSC	Head and Neck Squamous Cell Carcinoma
LUSC	Lung Squamous Cell Carcinoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach Adenocarcinoma
ER	Endoplasmic Reticulum
NB	Neuroblastomas
HCC	Hepatocellular Carcinoma
RTK	Receptor Tyrosine Kinase

BCSS	Breast Cancer Specific Survival
DRFS	Distanced Recurrence Free Survival
HR	Hazard Ratio
CI	Confidence Interval
NSCLC	Non-Small Cell Lung Cancer
ESCC	Esophageal Squamous Cell Carcinoma
CNA	Copy Number Alterations
ORF	Open Reading Frame
UTR	Untranslated Region
CDS	Coding Sequence
DMFS	Distance Metastasis Free Survival
PI	Propidium Iodide

List of figures

Figure 1. Stages of breast cancer development (Tower et al., 2019)	5
Figure 2. HER2 signaling pathway (Y. Feng et al., 2018)	7
Figure 3. A schematic showing the ways a fusion gene can occur at the chromosomal level (Mertens et al., 2015).....	10
Figure 4. Schematic representation of canonical biogenesis, processing and function of miRNA (Babashah & Soleimani, 2011).	13
Figure 5. VMP1-miR-21 locus (Z. Wang, 2013)VMP1-miR-21 and pri-miR-21 are two primary miR-21 transcripts.	16
Figure 6. Common metastatic sites in breast cancer	18
Figure 7. Breast cancer cell lines and tumors had 15 fusion genes in common.	44
Figure 8. The sequence of the junctions of fusion genes confirmed.	47
Figure 9. High VMP1 was marker of shorter overall survival.	49
Figure 10. VMP1 mRNA levels were higher in breast tumors than paired normal tissues.	50
Figure 11. VMP1 mRNA associated with VMP1 copy number variation.....	52
Figure 12. VMP1 mRNA expression associated with HER2 expression.....	54
Figure 13. VMP1 mRNA levels are highest in HER2-enriched and Luminal B subtypes.....	55
Figure 14. High expression of VMP1 was associated with shorter Breast Cancer Specific Survival (BCSS) in cohort 1 and METABRIC.....	57
Figure 15. High expression of VMP1 was associated with shorter DRFS time.....	58
Figure 16. VMP1 mRNA correlates with PPM1D, hsa-miR-21-3p and RPS6KB1	59
Figure 17. High expression of VMP1 was associated with shorter BCSS and DRFS.....	61

Figure 18. hsa-miR-21-3p levels were higher in breast tumors than paired normal tissues.	62
Figure 19. hsa-miR-21-is associated with MIR21 copy number variations.	63
Figure 20. Hsa-miR-21-3p was high in HER2 positive tumors.	67
Figure 21. Hsa-miR-21-3p was high in grade 3 tumors within cohort 2 and METABRIC.....	68
Figure 22. High expression of hsa-miR-21-3p was associated with shorter disease-free survival.	70
Figure 23. High expression of hsa- miR-21-3p was associated with shorter BCSS and DRFS within METABRIC/EGA.....	71
Figure 24. Levels of hsa-miR-21-3p were high only in ER ⁺ /HER2 ⁺ tumors.	73
Figure 25. In cultured cells, VMP1 mRNA levels were highest in ER+/HER2+ BC lines, consistent with data in BC tumors.	77
Figure 26. BT-474 cell lines express the most VMP1 protein.	79
Figure 27. Validation of VMP1 and ERBB2 knockdown.....	81
Figure 28. VMP1 silencing in BT474 cells did not affect proliferation.	82
Figure 29. VMP1 knockdown did not induce apoptosis.....	84
Figure 30. Silencing VMP1 did not affect ZO-1 expression.....	85
Figure 31. VMP1 silencing does not affect E-CAD expression.	86

List of tables

Table 1. PCR mix recipe for fusion genes confirmation.....	31
Table 2. Sequences of primers used for fusion-gene confirmation	32
Table 3. Primer sequences used for copy number variation analysis	34
Table 4. Thermal cycle conditions for copy number analysis	34
Table 5. RT-qPCR recipe.....	35
Table 6. q RT-PCR cycling conditions	36
Table 7. PCR mix recipe for miRNA quantification	37
Table 8. q RT-PCR cycling conditions for miRNA quantification	37
Table 9. SDS-PAGE sample buffer recipe.....	38
Table 10. List of primary and secondary antibodies	39
Table 11. Verified 15 common fusion with RT- PCR	45
Table 12. Five fusion genes passed filtering criteria [*]	46
Table 13. Amplification and correlation between DNA and mRNA of the gene partners that constitute the five fusion genes	48
Table 14. VMP1 CNV frequencies within cohorts.....	51
Table 15. Adjusted VMP1 cox model to RPS6KB1, PPM1D and hsa- miR-21-5p high groups	60
Table 16. hsa-miR-21-3p association with clinically pathologic features in METABRIC/EGA.	65
Table 17. Adjusted hsa- miR-21-3p cox model to neighbor genes expression and clinic-pathological features	72
Table 18. VMP1 expression status within HER2 positive BC cell lines	75

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-II.

- I. **Arsalan Amirfallah**, Adalgeir Arason, Hjorleifur Einarsson, Eydis T. Gudmundsdottir, Edda S. Freysteinsdottir, Kristrun Olafsdottir, Oskar Þor Johannsson, Bjarni A. Agnarsson, Rosa Bjork Barkardottir, Inga Reynisdottir, **High expression of the Vacuole Membrane Protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer**, published in (PLoS ONE 14(8): e0221413. <https://doi.org/10.1371/journal.pone.0221413>).
- II. **Arsalan Amirfallah**, Hildur Knutsdottir, Adalgeir Arason, Bylgja Hilmarisdottir, Oskar Thor Johannsson, Bjarni A. Agnarsson, Rosa Bjork Barkardottir, Inga Reynisdottir **Hsa-miR-21-3p associates with breast cancer patient survival and targets genes in tumor suppressive pathways** (submitted manuscript).

In addition, some unpublished data will be presented.

Declaration of contribution

The experiments described herein, including those done *in silico* and in “wet lab,” were performed by me, except for the quantification of hsa-miR21-3p and hsa-miR-21-5p in cohort 2 and breast cancer cell lines, which was performed by Guðrún Jóhannesdóttir (GJ) and Edda Sigríður Freysteinsdóttir (ESF). DNA and mRNA from tumors in cohort 1 and pathological and clinical data was from Adalgeir Arason, Bjarni A. Agnarsson (BAA), Oskar Th. Johannsson (OTJ) and Rosa B Barkardottir (RBB) research group in Landspítali. Bioinformatic analysis for submitted manuscript regarding hsa-miR-21-3p data was performed by Hildur Knutsdottir (HK).

The study was approved by The Icelandic Data Protection Commission (2001/523 and 2002/463) as well as the National Bioethics Committee of Iceland (99/051, 99/051_FS1, VSN11-105, VSN-15-138). DNA and mRNA isolation from samples in cohort 2 was performed on tissue from patients that had given informed consent to AA, BAA, OTJ and RBB for scientific studies on breast cancer (BC). The National Bioethics Committee gave the research group permission to use their research material in collaboration projects with Inga Reynisdottir (IR) on her studies on BC. The DNA and RNA isolation from tissues of patients in cohort 2 was mostly performed by ESF and GJ and a few other, previous employees at the Laboratory of Molecular Pathology and Cell biology. All experiments were designed by me and my supervisor IR. Hypotheses were generated, and results interpreted done by me and IR. Paper 1 was prepared by me and written by me and IR. The second manuscript is in preparation by IR and me.

1 Introduction

1.1 Epidemiology

According to the World Health Organization's statistics on cancer, for year 2018, the worldwide frequency of breast cancer, among all cancer diagnoses in women, was 24.2%, with an overall 6.6% mortality rate. For the US in 2020, statistical estimates show 30% of female cancers are breast cancer and it is the second most common cause of death due to cancer (Siegel, Miller, & Jemal, 2020). In an Icelandic report of cancer registry statistics for years 2014 to 2018, 27% of all cancer types were breast cancer, with a 14.96% mortality, on average, per year (Krabbameinsskrá).

1.2 Etiology

Breast cancer has a complicated etiology, with both genetic and non-genetic factors influencing its development (Britt, Cuzick, & Phillips, 2020). Alongside with genetic and non-genetic factors, cellular factors such as , metabolic intermediates, miRNA and signaling molecules and interactions with stromal cells also have roles in etiology of cancer development including breast cancer (Paul, 2020).

Genetic factors like BRCA1 and BRCA2 (DNA repair genes) are the most common breast cancer susceptibility genes; and women with BRCA1 and BRCA2 gene mutations have 72% and 69%, respectively, greater risk of getting breast cancer by age 80 (Kuchenbaecker et al., 2017). In addition, rare germline mutations in CDH1, PTEN, STK11, TP53, CHEK2, ATM, NBN, and PALB2 genes carry risk for breast cancer development (Easton et al., 2015). Low-penetrance SNPs, typically located in non-coding regions of the genome, have noteworthy effects on breast cancer etiology and predict pathological subtypes, supporting the idea that breast cancer subtypes arise through distinct etiological pathways. These genes appear to be important in breast tumor development and interact with environmental and hereditary factors (Broeks et al., 2011; Michailidou et al., 2015; Suvanto et al., 2020). Accumulation of somatic mutations over the time leads to formation of cancers (Jolly & Van Loo, 2018); and in large-scale sequencing projects of breast tumors mutations were found to associate with particular molecular, environmental, and endogenous exposures (Bodily et al., 2020).

Non-genetic factors like obesity and alcohol consumption have been shown to increase risk of getting ER positive breast cancer (N Hamajima, 2019), younger menarche age, older age of menopause, having fewer children and at an older age, shorter breastfeeding periods, mammographic density, and physical inactivity, are all non-genetic factors that increase breast cancer risk. Having a healthy life style with physical activity, reduced fat intake, and increased consumption of vegetables and grains all can reduce the breast cancer risk (Britt et al., 2020).

1.3 Clinical pathology

Effective patient management requires an understanding of all clinically pathological features of breast cancer, such as tumor type, histograde type, tumor size, lymph node involvement and receptor status (i.e., ER, PGR and HER2) (Harbeck et al., 2019).

1.3.1 Histograde

Breast cancer has 20 different histological subtypes classified according to morphology and growth pattern. The Invasive Ductal Carcinoma (IDC) subtype accounts for 80% of breast cancer, with the remaining being Invasive Lobular Carcinoma (ILC), medullar, metaplastic, tubular, and mucinous subtypes, each classified as special histological types with a distinctive growth pattern and variable prognosis (Lakhani SR, 2012). The degree of differentiation, mitotic index, or proliferative activity and aggressiveness are parameters for grading breast tumors. As such, the Nottingham Grading System grades tumors according to a three-tiered scoring system (Rakha, Reis-Filho, Baehner, et al., 2010).

1.3.2 Staging (TNM)

Based on the American Joint Committee on Cancer (AJCC) TNM staging system for breast cancer, T refers to the size of tumor and its invasion into the chest wall (<2 cm, between 2 to 5 cm and >5 cm), N measures the number of lymph nodes with cancer (0, 1–3, 4–9, >10), and M measures the distance of metastasis (Giuliano, Edge, & Hortobagyi, 2018).

1.3.3 Receptors

Estrogen Receptor (ER), Progesterone Receptor (PR) and Human Epidermal Growth Factor Receptor 2 (HER2/ERBB2) are key diagnostic biomarkers in breast cancer. Expression of ER, PR, and HER2 proteins is assessed with immunohistochemistry in breast tumors, and correlates with behavior of

tumors, patient outcome, and response to endocrine therapy or HER2-targeted therapy (Rakha, Reis-Filho, & Ellis, 2010). Nearly 75% of all breast cancer patients are ER⁺, 55 to 65% are PR⁺, and 13 to 20% are HER2 positive or ERBB2 amplified. ER⁺ and PR⁺ breast tumors are named as Hormone Receptor positive (HR⁺). HR positive tumors receive endocrine therapy, at the same time HER2⁺ breast tumors are treated with trastuzumab or other HER2-targeted therapies (Howlader et al., 2014). Aromatase inhibitors (AIs) are used in treatment of HR positive postmenopausal breast cancer patients; and they are much more effective when used as adjuvant therapy to chemotherapy and surgery. HR positive BC is less aggressive (compared to HR negative BC and HER2⁺ or ERBB2 amplified BC) and has better prognosis due to benefits from endocrine therapy (Y. Li et al., 2020).

1.3.4 Molecular subtypes

Sørli et al. used RNA microarray gene expression analysis to classify breast cancer among five intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, Basal and Normal-like (Sørli et al., 2001). Luminal subtypes are ER⁺ and based on Ki67 expression; they are divided into two subgroups, Luminal A and Luminal B. Luminal A subgroup tumors are ER/PR positive, HER2 negative, with low expression of Ki67, a low grade, and with good prognosis. Luminal B tumors may be ER/PR / HER2 negative or positive with high expression of Ki67, a high grade, and with a worse prognosis (Provenzano, Ulaner, & Chin, 2018). HER2-enriched subtype tumors encompass those with amplification of ERBB2 and overexpression of HER2 protein; they are ER/PR negative, grow faster than Luminal B tumors, and had a worse prognosis before use of HER2-targeted therapies. Basal-Like/Triple-Negative subtype tumors are ER, PR, and HER2 negative, with high expression of Ki67, a high grade, and a poor prognosis. The basal-like subgroup is the most common in women with the BRCA1 mutation, and in young, and African American women. The basal-like and triple negative breast cancer are not synonymous. Classification of basal-like is done according to expression pattern of many genes such as ER, PGR and ERBB2 genes. Triple negative breast tumors classification is done according to immunohistochemical staining of ER, PGR and HER2 proteins and do not express ER, PR and HER2 proteins. Triple negative tumors include several subtypes and basal-like tumors composes sizable portion of them (Y. M. Lee, Oh, Go, Han, & Choi, 2020). The Normal-like subtype has a gene expression pattern similar to normal breast tissue and mostly consists of adipose tissue and, like the Luminal A subtype, is ER or PR positive, HER2 negative, Ki67 negative, with an intermediate prognosis (Dai et al., 2015).

1.3.5 Breast cancer diagnosis

Appropriate diagnostic evaluations must be made when women experience breast changes and symptoms, such as a lump, or pain in their nipples and breasts. Clinical diagnosis of breast cancer is based on a result of triple tests, including clinical examination, imaging, and needle biopsy. Elements of the triple test should be performed before treatment to discriminate cancer from benign conditions and normal changes. Clinical examination consists of palpation of breasts and regional lymph nodes, and distant metastasis assessment. Imaging of breast cancer includes of “bilateral mammography.” MRI is not recommended for routine diagnosis of breast cancer unless in cases of familial breast cancer in connection with BRCA mutations, lobular subtype, or to facilitate imaging of dense breasts or evaluating response to therapy and before neoadjuvant systemic therapy (Cardoso et al., 2019).

1.3.6 Therapy

The therapeutic goals for treating non-metastatic breast cancer are eradication of tumor cells from breast and regional nodes and preventing disease progression. Therapy of non-metastatic breast cancer includes surgery, resection of axillary lymph nodes, preoperative or postoperative radiation, and preoperative (neoadjuvant) or postoperative (adjuvant), or both systemic therapies. Standard systemic therapy is administered based on subtype of breast cancer. For treatment of triple-negative breast cancer, chemotherapy alone or with immunotherapy is used; for treatment of Hormonal receptor (HR) positive tumors, hormonal treatment alone or with chemotherapy is used; and for treatment of HER2 positive tumors, HER2-directed therapy with chemotherapy is used and if they are HR positive, endocrine therapy is given as well. The treatment strategy for metastatic breast cancer is similar to the treatment for non-metastatic breast cancer, with the goals of prolonging patient lifespan and palliation of symptoms (Waks & Winer, 2019). Targeted therapy includes aromatase inhibitors for treatment of estrogen dependent tumors, recombinant and conjugated monoclonal antibodies for HER2 positive tumors, and PARP1 inhibitors for treatment of triple-negative tumors with BRCA1 or BRCA2 germline mutations (Masoud & Pagès, 2017). Immunotherapy is used in 1% to 2% of breast tumors, with alterations in the mismatch repair pathway and high microsatellite instability (Cortes-Ciriano, Lee, Park, Kim, & Park, 2017).

1.4 Tumor biology

1.4.1 Development

Carcinoma of the breast tissue can originate from epithelium of ducts or lobules before spreading to the rest of the body, in what is termed a pre-invasive lesion. In pre-invasive lesions, cancer cells are confined to the ducts or lobules and have not yet broken through the basement membrane. The pre-invasive lesions in ducts are categorized as atypical ductal hyperplasia (ADH) when they have initiated and expanded within the ducts, ductal carcinoma *in situ* (DCIS) when they have filled ducts with tumor cells, and invasive ductal carcinoma (IDC) once the tumor cells break through the basement membrane and invade the surrounding stroma (Malhotra, Zhao, Band, & Band, 2010; Tower, Ruppert, & Britt, 2019)

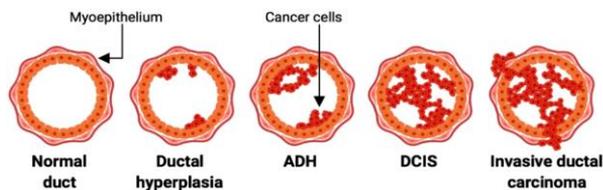


Figure 1. Stages of breast cancer development (Tower et al., 2019)

Initiation and expansion of tumor cells within atypical ductal hyperplasia (ADH), Progress of ADH to ductal carcinoma in situ (DCIS) defines as complete filling of mammary ducts with tumor cells.

1.4.2 Altered signaling pathways in breast cancer development

Genetic and epigenetic alterations of signaling pathways in cancer cells lead to activation of proto-oncogenes and inactivation of tumor suppressor genes, which causes uncontrolled cell proliferation and division of cancer cells. Estrogen receptor (ER) signaling and HER2 signaling pathways are among the predominant signaling pathways in breast cancer development and progression (Y. Feng et al., 2018).

1.4.2.1 ER signaling

Both nuclear Estrogen Receptor alpha and beta (ER α , ER β) are transcription factors with common structural features and the ability to heterodimerize; they are encoded by the ESR1 and ESR2 genes, respectively. In breast cancers, ER α expression is higher than ER β ; and ER α plays a critical role in pathogenesis, since ER α interacts with cyclin D1 and promotes development of breast cancer. The ER α /cyclin D1 interaction activates cyclin-dependent

kinases (CDKs) 4 and 6, which promotes the G1 to S phase transition of the cell cycle, thereby driving proliferation of malignant breast cells (Y. Feng et al., 2018; Renoir, Marsaud, & Lazennec, 2013).

1.4.2.2 HER2 signaling

Human Epidermal growth factor Receptor-2(HER2) is one of four members of the epidermal growth factor receptor (EGFR) family and is a receptor tyrosine kinase with 3 domains, including an extracellular ligand-binding domain, a transmembrane domain, and an intracellular domain (Roskoski, 2014). Ligand binding causes HER2 dimerization and phosphorylation of tyrosine residues in the HER2 intracellular domain, which activates oncogenic signaling pathways including mitogen-activated protein kinase (MAPK) and the phosphatidylinositol 4,5-bisphosphate 3-kinase (PI3K) both heavily associated with breast cancer development (Figure 2)(Wee & Wang, 2017). Amplification of ERBB2 in breast tumors (King, Kraus, & Aaronson, 1985) results in HER2 overexpression and cancer development and progression. HER2 positive tumors are aggressive and more prone to progression and metastasis than HER2 negative ones. The HER2 signaling pathway is the target of various drugs such as monoclonal antibody like trastuzumab, dual tyrosine kinase inhibitor like lapatinib and conjugated trastuzumab ado-trastuzumab emtansine, and patients who respond to these medications have a good prognosis. Clinically, routine diagnostic tests assessing ERBB2 amplification and HER2 protein expression are used to select patients for HER2 targeted therapy, which has had an undeniable benefit for improving survival of breast cancer patients (Nwabo Kamdje et al., 2014).

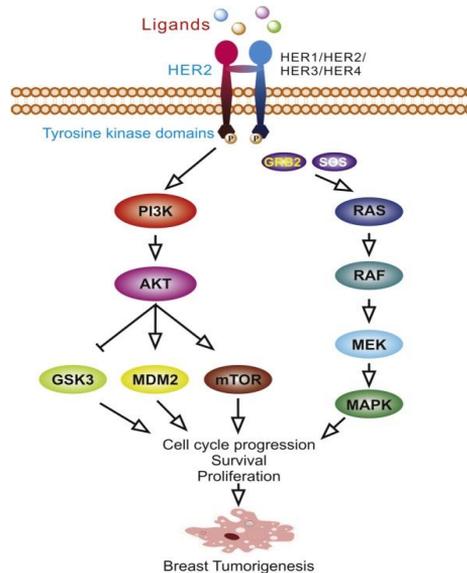


Figure 2. HER2 signaling pathway (Y. Feng et al., 2018)

HER2 is a receptor tyrosine kinase, located on the cell membrane. Variety of ligands activate HER2. Activation of HER2 in response to ligands leads to phosphorylation of its tyrosine kinase domain in the cytoplasm and subsequent initiation of downstream oncogenic signaling pathways such as PI3K/AKT pathway and Ras/MAPK pathway.

1.4.3 Known gene mutations in breast cancer

Gene mutations in breast cancer, like with other cancers, activate protooncogenes and often lead to uncontrolled and continuous activity of mutated proteins (Kufe DW, 2003). Genetic alteration of both the germline and somatic genome are common causes of breast cancer (Ramroop, Gerber, & Toland, 2019). Somatic mutation refers to DNA alterations in of cells other than germline cells (sperm and egg), and includes point mutations, single nucleotide variants, somatic indels, and copy number alterations (Little, Lin, & Sun, 2019). Germline mutations refer to alterations in DNA of reproductive cells (egg and sperms) and are inherited from parents to offspring. They include single base pair deletions, insertions, duplications, and often introduce amino acid changes (Yadav & Couch, 2019).

Germline mutations in the breast cancer susceptibility genes 1 or 2 (BRCA1 or BRCA2) were identified in 20% of women with a history of breast cancer in their first-degree relatives. The BRCA proteins encoded by BRCA1 and BRCA2 genes possess strong tumor suppressing ability and participate in repairing DNA damage through homology-directed repair (HDR). Thus,

loss-of-function BRCA mutations decrease DNA repair efficiency and increase the risk of breast cancer development by six-fold (Kuchenbaecker et al., 2017). Mutations in BRCA1 and BRCA2 genes guide therapy selection in breast and ovarian cancer patients and predict responsiveness to platinum-based and of poly (ADP-ribose) polymerase (PARP)-inhibitor-based chemotherapies (Tung & Garber, 2018).

The phosphoinositide 3-kinase (PI3K) signaling pathway has a key role in breast cancer development and is altered in hormone receptor positive tumors. Somatic mutations of the catalytic subunit alpha of PI3K (PIK3CA) can activate the PI3K pathway, triggering cell proliferation, resistance to apoptosis, and increases in cell-cycle progression and translation. The frequency of mutation in PIK3CA, across breast cancer subtypes, is between 20% to 45%, and has a prognostic value for the PI3K targeted therapies (Mollon et al., 2020).

The tumor protein 53 (TP53) is a well-known tumor suppressor and is frequently mutated in all cancers including breast cancer. Overall, TP53 is mutated in 30 to 35% of primary breast cancers, depending on the breast cancer molecular subtype. Mutation in TP53 is highest in triple negative and is lowest in Luminal A breast tumors and has been shown to be elevated in early tumorigenesis, and during tumor growth, development, and metastasis (Duffy, Synnott, & Crown, 2018). Unlike somatic mutations, the frequency of germline mutations of the TP53 gene in breast cancer is extremely low (Walerych, Napoli, Collavin, & Del Sal, 2012). Phosphatase and tensin homolog (PTEN) is a tumor suppressor gene and one of the most frequently altered genes in breast cancer. It plays a role in cell-cycle progression, cell growth, and survival of cancer cells. The 3'-group of the phosphatidylinositol (3, 4, 5)-trisphosphate (PIP3) is dephosphorylated by PTEN, leading to inactivation of the PI3K/Akt pathway and inhibition of growing of tumor cells. PTEN also has a role in binding to an increasing TP53 stability, DNA repair, and genome stability. Nearly 40% to 50% of breast tumors have lost of heterozygosity at the PTEN locus. Inactivation of PTEN through somatic mutation, or mono and bi allelic loss leads to over activation of PI3K pathway and proliferation of cancer cells (Bazzichetto et al., 2019; Carbognin, Miglietta, Paris, & Dieci, 2019).

1.5 Chromosomal rearrangements in breast cancer

Chromosomal rearrangements are one of the mechanisms of proto-oncogene activation in all cancers, including breast cancer, and lead to structural changes of the genome caused by DNA breakage, and incorrect rejoining and replication (Paratala et al., 2016). Chromosomal rearrangements are frequent in breast tumors and among these are copy number alterations (CNA) and gene fusions (Mertens, Johansson, Fioretos, & Mitelman, 2015) .

1.5.1 Copy number alterations (CNA)

Copy number alterations (CNAs) are somatic changes in the number of copies of genomic material, in the form of either gain or loss (Zhang, Feizi, Chi, & Hu, 2018). The mechanism of copy number gain and loss can activate oncogenes, alter of tumor suppressor gene copy number, and lead to tumor development. Thus, they can be therapeutic targets. An example is Trastuzumab for breast cancer patients carrying ERBB2 amplification (Haverty et al., 2008). Expression of 85% of genes in breast cancer, are due to somatic copy number variation at gene loci. These genes, expressed due to somatic copy number variation, are often oncogenes or tumor suppressor genes that have a direct effect on disease progression (Zhang et al., 2018). Gene amplification frequently occurs in most cancer types, including breast cancer; and amplification of HER2, EGFR, MYC, CCND1, MDM2, AIB1, FGFR1, S6K, TOPO2A, EMS1, FGF3, AKT2, and PIP4K2 genes has been confirmed through various studies in breast tumors (Al-Kuraya et al., 2004). The 8p12, 8q24, 11q13, 17q12 , 20q13 and 12p13 chromosome regions are frequently amplified in breast tumors, and each region contains important target genes (Yao et al., 2006). There are various chromosomal regions like 1p, 1q, 3q, 4q, 6p, 6q, 8p, 10p, 14q, 15q, 16p, and 19q found with amplification but without identified target genes. Since DNA is more stable than RNA and protein, measuring amplification is easier to use as diagnostic analysis, like ERBB2 amplification. Gene amplification can also reflect increased genetic instability and is an indicator of poor prognosis in breast cancer patients (Al-Kuraya et al., 2004).

Copy number alterations are well characterized in breast cancer (Silva et al., 2015). Recurrent CNA have been identified using array-based technologies; and recently, to overcome intra-tumoral heterogeneity, single-cell, DNA-sequencing methods are being applied. Loss of the X-chromosome and its association with ER negative breast tumors, as well as the detection of pseudo-diploid cells, are the result of using of modern technologies to analyze CNA (Baslan et al., 2020). Based on whole-genome sequences data from primary breast tumors, recurrent copy number changes (like amplification and homozygous deletions) generate driver mutations in cancer genes (Nik-Zainal, 2016).

1.5.2 Fusion genes

A fusion gene is a hybrid gene formed by chromosomal rearrangement of two previously separate genes. Gene fusions are found at a significantly higher rate in cancer samples compared to benign samples and play critical roles in carcinogenesis via various mechanisms such as oncogene activation, tumor suppressor deletion/downregulation, and the creation of novel proteins capable of altering cellular pathways (Paratala et al., 2016).

In hematologic malignancies, the Philadelphia chromosome results in the BCR-ABL fusion, causing inappropriate activation of ABL kinase; this was the first reported fusion genes, fifty years ago. Fusion genes can be diagnostic markers and direct targeted therapy with Imatinib has been used to treat this type of cancer for years. The tumorigenic properties of fusion genes, in most cancer types including breast cancer, has recently attracted attention (Paratala et al., 2016). Balanced and unbalanced chromosome rearrangements may lead to formation of fusion genes. Balanced changes include translocations, insertions and inversions, where translocation refers to “the transfer of chromosome segments between chromosomes”, insertion to a new interstitial position of a chromosome segment in the same or another chromosome, and inversion refers to a double break of and internal chromosome segment that flips 180 degrees and rejoins the chromosome. Unbalanced chromosomal changes include “deletion of an interstitial chromosomal segment” (figure 3). Both balanced and unbalanced chromosomal rearrangements can lead to deregulation of gene A or gene B, or formation of chimeric or truncated gene (Mertens et al., 2015)

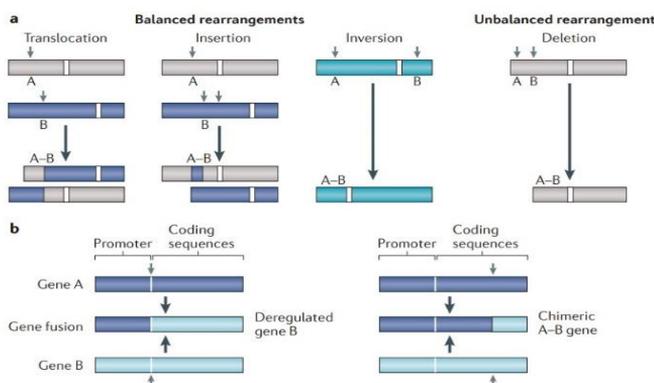


Figure 3. A schematic showing the ways a fusion gene can occur at the chromosomal level (Mertens et al., 2015).

Balanced and unbalanced chromosomal rearrangements lead to generation of fusion genes. Balanced changes comprise translocations, insertions, and inversions; deletion of an interstitial chromosomal segment is an example of an unbalanced change that leads to fusion genes generation.

Gene fusions mostly result from insertion, deletion, inversion, or tandem duplication or amplification, and may involve the same chromosome (intra-chromosomal) or different chromosomes (inter-chromosomal). According to multiple whole-genome sequencing studies, the majority of chromosomal rearrangements are associated with intra-chromosomal tandem duplications and amplifications (Kumar-Sinha, Kalyana-Sundaram, & Chinnaiyan, 2015).

1.5.3 Known fusion genes in breast cancer

Gene fusions recurrently found in breast cancer are rare in general, but in recent years with the improvement in sequencing technologies and fusion-finding bioinformatics pipelines, several recurrent and pathological fusion genes were identified in aggressive subtypes like luminal B, triple negative, endocrine resistance breast cancer and secretory breast cancer (Nik-Zainal, 2016). Fusion genes with open reading frames (ORFs) are rare breast tumors and transcriptomic sequence of ORF fusion genes demonstrated low expression of them in breast tumors, still they can abrogate expression of the participating genes (A + B)(Nik-Zainal, 2016) ETV6-NTRK3 (Tognon et al., 2002) and MYB-NFIB (M. Persson et al., 2009) fusions in secretory breast cancer, ESR1-CCDC170 fusion genes in ER positive luminal B and endocrine resistance subtypes (Veeraraghavan et al., 2014), SCNN1A-TNFRSF1A and CTSD-IFITM10 fusion genes in triple negative and ER positive subtypes (Varley et al., 2014), and other recurrent gene fusions (like RPS6KB1-VMP1 (Inaki et al., 2011), EEF1D3-FRY (Kim et al., 2015; Nik-Zainal, 2016) and NOTCH1 and NOTCH2 family rearrangements (Robinson et al., 2011) in mixed-subtype breast cancers, are all examples of recurrent fusions in some aggressive and mixed subtypes of breast cancer.

Currently, Entrectinib (an oral inhibitor of the tyrosine kinases, NTRKs, ROS, and ALK), is being tested in clinical trials (NCT02097810& NCT02568267) for breast cancer patients with ETV6-NTRK3 gene fusions (Veeraraghavan et al., 2014). RPS6KB1-VMP1 is generated by tandem duplication and recently reported in 30% of breast cancer patients from Singapore. Expression levels of this fusion is very low in normal breast tissues; and the chimeric fusion protein does not produce a functional protein and is not a driver of tumor development (Inaki et al., 2011) . According to the most recent study, additional fusions found VMP1 involved as a 3' fusion in breast tumors with CLTC/ and AC099850 and VMP1 as a 5'-end gene in HER2 positive tumors (H. Persson et al., 2017). In addition, to the above mentioned breast cancer fusion genes, other recurrent fusion genes include ERLIN2-FGFR1, FGFR2-AFF3, FGFR2-CASP7, and FGFR2-CCDC6, where

FGFR family member, fused at the 3' or 5' end to their partner, were shown to trigger the activation kinase domain of FGFR (Y. M. Wu et al., 2013).

1.6 miRNAs

1.6.1 History

Discovery of miRNAs in 1993 by Lee et al. in *Caenorhabditis elegans* (R. C. Lee, Feinbaum, & Ambros, 1993) shed light on their role in regulating expression of mRNAs and had substantial impact on new discoveries regarding gene regulation in the years that followed (Bertoli, Cava, & Castiglioni, 2015). miRNA sequence identification and annotation was first attempted in 2006, with only 218 miRNA loci and assembled into the miRBase (Griffiths-Jones, Grocock, van Dongen, Bateman, & Enright, 2006). The miRBase dataset currently contains 1,917 miRNA entries for the human genome (Harbeck et al., 2019).

1.6.2 Biogenesis

miRNAs can be transcribed from individual genes with their own promoter or generated as transcripts inside other protein-coding genes, making RNA polymerase II a key enzyme in facilitation of miRNA genes transcription. During processing of pri-miRNA by Drosha and double-stranded RNA binding protein DGCR8, pre-miRNA forms a characteristic stem-loop precursor and is exported to the cytoplasm because of interaction of exportin-5 and Ran- GTP with pre-miRNA. In the cytoplasm, dicer removes the pre-miRNAs loop structure and produces a duplex molecule consisting of the single stranded mature miRNA, miR-3p, and miR-5p fragments. A helicase unwinds the miR-3p: miR-5p duplex, resulting in degradation of miR-3p (i.e., the passenger strand). Next, the mature miRNA and miR-5p (the guide strand) bind the Argonaut (Ago) protein and are thereupon incorporated into RISC. The RISC-miRNA complex can target mRNA for degradation, thereby repressing translation (Figure 4)(Babashah & Soleimani, 2011; Bertoli et al., 2015; Czech & Hannon, 2011; Stavast & Erkeland, 2019).

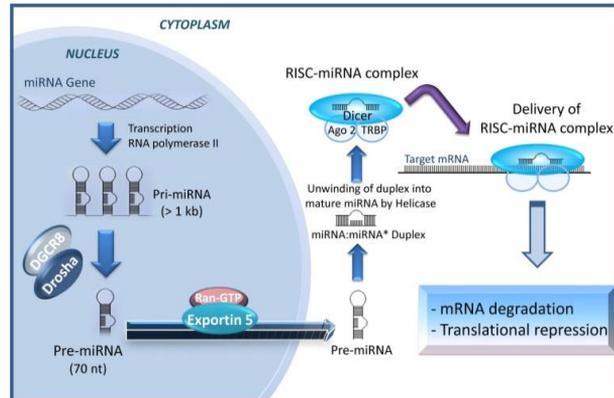


Figure 4. Schematic representation of canonical biogenesis, processing and function of miRNA (Babashah & Soleimani, 2011).

RNA polymerase II primarily facilitates transcription of the miRNA gene in the nucleus. The resultant pri-miRNA transcript is processed by Drosha, producing a characteristic stem loop precursor, pre-miRNA. The pre-miRNA is then exported into the cytoplasm by exportin-5 and Ran-GTP. In the cytoplasm, final processing mediated by Dicer removes loop structures of pre-miRNAs, producing a duplex molecule containing the single stranded mature miRNA and a miRNA* fragment. The miRNA:miRNA* duplex is unwound by Helicase; the miRNA* fragment is degraded, whereas the mature miRNA molecule binds to an Argonaute (Ago) protein and incorporates into the RISC. The RISC-miRNA complex can then target mRNAs bearing a perfectly complementary target site for degradation or can repress the translation of an mRNA that shows imperfect complementarity with the small RNA. Primary miRNA, pri-miRNA; precursor miRNA, pre-miRNA; Drosha, RNase III endonuclease; DGCR8, DiGeorge syndrome critical region 8; Dicer, RNase III endonuclease; RISC, RNA-induced silencing complex.

1.6.3 Role of miRNAs in cancer

Micro RNA expression is generally lower in tumor cells than normal cells; and in normal cells regulates apoptosis, proliferation, differentiation, cellular growth, and cell-cycle progression. microRNAs may play key roles in tumorigenesis (e.g., regulation of tumor suppressors or oncogenes) making them novel cancer therapy targets. microRNAs are linked to genetic dysregulation and epigenetic changes like amplification, deletion and point mutations. Notably, most microRNAs reside on fragile, amplified regions of chromosomes (Bertoli et al., 2015; Hemmatzadeh, Mohammadi, Jadidi-Niaragh, Asghari, & Yousefi, 2016; O'Day & Lal, 2010).

1.6.4 Known miRNAs in breast cancer

Profiling of microRNAs through various studies in breast cancer has led to

identification of microRNAs, which like other cancer genes can function as tumor suppressors and oncogenes, promoting or inhibiting metastasis and correlating with breast tumor progression (Klinge, 2018; O'Day & Lal, 2010). The most well-known breast cancer microRNAs are discussed in the following section.

1.6.4.1 Tumor suppressor miRNAs

1.6.4.1.1 miR-206

miR-206 is downregulated in breast tumors compared to normal breast tissues (Iorio et al., 2005) and in metastatic breast cancer cells compared to primary tumors (Tavazoie et al., 2008), and likely suppresses breast cancer tumorigenesis. Upregulation of miR-206 in estrogen negative breast tumors pointed to its role in regulation of ESR1 gene. Accordingly, this microRNA was shown to inhibit expression of ESR1 through binding sites in the 3' UTR of ESR1 mRNA (Adams, Furneaux, & White, 2007).

1.6.4.1.2 miR-17-5p

miR-17-5p (or miR-91) lies in the chromosome 13q31 region, a region that frequently loses heterozygosity in breast tumors as well as many other cancer types (Eiriksdottir et al., 1998). Low Expression of miR-17-5p can be a predictive marker for recurrent breast tumors (Y. Wang et al., 2018). Proliferation of primary breast tumors is inhibited with miR-17-5p because it targets oncogenes like AIB1, CCND1, and transcription factors like ER α and E2F1 (Hossain, Kuo, & Saunders, 2006; Z. Yu et al., 2008).

1.6.4.1.3 miR-200 family

miR-200 induces the 'epithelial-mesenchymal transition' (EMT) by suppressing the EMT inducers ZEB1 and ZEB2 is an important biomarker for breast cancer progression and metastasis. In MDCK cells, inhibition of miR-200a, b, and c induces the EMT, increasing expression of vimentin, fibronectin, and N-cadherin, and decreasing expression of E cadherin; and the expression profile was predictive of the metaplastic breast tumor type, a highly aggressive phenotype (Gregory et al., 2008; O'Day & Lal, 2010).

1.6.4.1.4 let-7 family

The let-7 family first was shown to have role in development of *C. elegans*. They are underexpressed and often deleted in various cancers including breast cancer. In vivo overexpression of let-7 family miRNAs in breast tumors

initiating cells (BT-ICs) reduced proliferation, and the tumor-forming and metastatic capacity of these cells (Q. Guo et al., 2018; F. Yu et al., 2007).

1.6.4.1.5 miR-34a

In multiple cancers, miR-34a is downregulated (Gaur et al., 2007) miR-34a is regulated by P53, and, in the context of breast cancer, is expressed less in triple-negative than in HER2-positive and mesenchymal lines (Kato et al., 2009). It was shown that pathological activation of the p53 network leads to decreased expression of miR-34 family (Imani, Wu, & Fu, 2018).

1.6.4.1.6 miR-31

miR-31 is abundantly expressed in normal breast cells and suppresses metastasis by inhibiting pro-metastatic genes. This anti-metastatic effect makes miR-31 a likely therapeutic target in the future (Luo et al., 2016; Valastyan et al., 2009).

1.6.4.2 Oncogenic miRNAs

1.6.4.2.1 miR-21

Overexpression of miR-21 is characteristic of human cancers, including breast cancer (O'Day & Lal, 2010). One of the first reported, miR-21 (Iorio et al., 2005) is abundantly expressed in breast tumor tissues compared to normal tissues., and has been linked to the patient's survival, stage, and node and metastasis status (Chen & Wang, 2014). Programmed cell death 4 gene (PDCD4)(Frankel et al., 2008) and several well-known tumor suppressor genes—tropomyosin 1 (TPM1)(Zhu, Si, Wu, & Mo, 2007), phosphatase and tensin homolog (PTEN), maspin, chromosome condensation protein G (NCAPG), and reticulon four isoform A (RTN4)—are all cancer-related targets of miR-21 (Zhu et al., 2008). As such, it is a diagnostic marker for early detection of breast cancer. Notably, the sequence of both stands of miR-21 (hsa-miR-21-3p & hsa-miR-21-5p) is highly conserved within species (Selcuklu, Donoghue, & Spillane, 2009).

miR-21 is located at the 17q23.2 locus, within the coding region of the vacuole membrane protein 1 (VMP1) gene. VMP1 and miR-21 have their own promoters but overlap for ~4kb of the genome and share 1kb of expressed RNA sequence. This chromosomal locus is complex, and often the site of rearrangement. Primary miR-21 (primi-21) transcription starts within intron 10th of the VMP1 gene, but the transcripts use different polyadenylation signals. VMP1-miR-21 (spliced) and pri-miR-21 (non-spliced)

are 2 transcripts of pri-miR-21 (Figure 5). Non-spliced pri-miR-21 transcripts terminate at their own poly adenylation signal sequence, downstream of the miR-21 hairpin. Spliced VMP1-miR-21 transcripts arise from the coding region of the VMP1 gene (Z. Wang, 2013).

In cancerous tissues, AP-1 induces miR-21 expression and AP-1 activation can induce both VMP1 and VMP1-mir21 (Fujita et al., 2008). In cancerous and inflamed tissues, the IL-6/STAT3 pathway contributes to overexpression of miR-21 (Ribas & Lupold, 2010). There are two CpG islands located in the VMP1-miR21 locus, one near to promoter of miPPR-21 and a second upstream of the start site of VMP1 transcription. Only the miPPR-21--associated CpG island was shown to be methylated (Ribas et al., 2012). Cancer-related pathways, like DNA repair factor MSH2, the metabolic enzymes CYP27B1 and PPARa, Tropomyosin alpha-1 chain (TPM1), Poly(rC)-binding protein 1 (PCBP1), and Tissue Plasminogen Activator (PLAT) are all reported to be targets of miR-21 (Z. Wang, 2013).

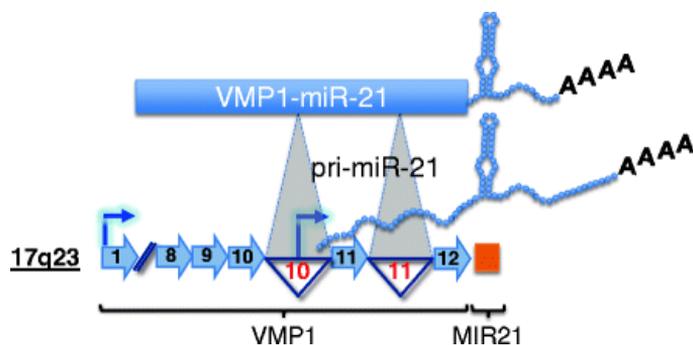


Figure 5. VMP1-miR-21 locus (Z. Wang, 2013) VMP1-miR-21 and pri-miR-21 are two primary miR-21 transcripts.

Terminal poly(A) tails are labeled as "A." Coding exons are depicted as light blue arrows whereas the two last intronic regions of VMP1 are highlighted as inverted triangles. Promoters and transcription start sites are indicated as dark blue arrows. The orange square represents the genetic location of the miR-21 hairpin.

1.6.4.2.2 miR-10b

miR-10b is a metastasis-specific micro-RNA with a role in the metastatic potential of cancer cells. In breast tumors, it is only expressed in metastatic cells. Based on both *in vitro* and *in vivo* experiments, miR-10b promotes migration and invasion of tumor cells. The transcription factor Twist induces miR-10b expression (Ma, Teruya-Feldstein, & Weinberg, 2007).

1.6.4.2.3 miR-373/520c family

Based on *in vitro* and *in vivo* experiments, miR-373 and miR-520c are known to have roles in promoting cancer cell migration and invasion. miR-373 is highly expressed in metastatic breast tumors that also express little CD44 (O'Day & Lal, 2010).

1.7 Breast cancer recurrence:

In 30% of patients and up to 6 to 10 years after the initial diagnosis, breast cancer can re-emerge—fundamental problem for treatment (Anandan, Sharifi, & O'Regan, 2020). Recurrence can be loco-regional or distant. Time of breast cancer recurrence can be influenced by adjuvant treatment strategies and classic prognostic factors, like ER and HER2 status. Breast cancer cells mostly spread to lymph nodes, brain, bone, lungs, and liver (Figure 6). Metastasized breast cancer can be treated to relieve symptoms and prolong life expectancy but is virtually incurable (Colleoni et al., 2016; Harbeck et al., 2019; Holleczeck, Stegmaier, Radosa, Solomayer, & Brenner, 2019). Young age at diagnosis, large tumor size, high grade, vascular invasion, and regional lymph node involvement are factors that can influence loco-regional metastasis (Holleczek et al., 2019). The expression of HER2 and estrogen receptor (ER) is a key aspect for increased risk of distant metastasis to specific organs. Early recurrence mostly happens in ER- tumors whereas ER+ tumors carry a risk of recurrence even five years later (Xiao et al., 2018). HER2 positive and triple negative subtypes mostly metastasize to the brain (Xiao et al., 2018). Age and comorbidities of other chronic diseases of cancer patients influence recurrence and understanding patterns of recurrence and time should improve targeted therapy approaches and patient outcomes (Colleoni et al., 2016). Alongside with above mentioned factors, tumor cells dormancy and reactivation have prominent roles in invasion and metastasis. Cancer cell dormancy occurs during the primary tumor formation phase or after invasion to secondary organs (Fares, Fares, Khachfe, Salhab, & Fares, 2020).

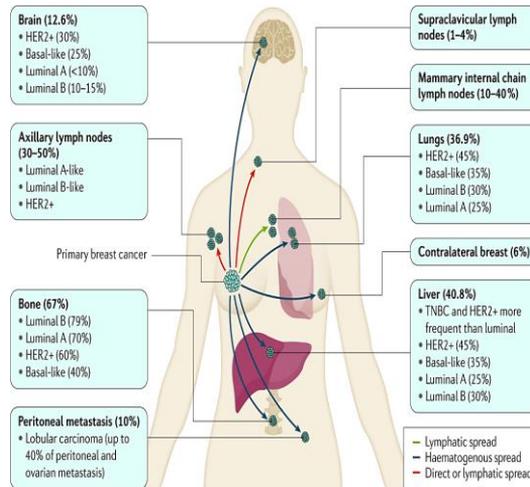


Figure 6. Common metastatic sites in breast cancer (Harbeck et al., 2019). Axillary lymph nodes are the most frequent nodal site breast tumors involve and internal mammary chain is another site for 10–40% of breast cancers to metastasis. Distant metastasis of breast tumors influenced with molecular subtypes. Common distant metastatic sites are brain, lungs, bone, and liver.

1.7.1 Pathophysiology of metastasis

The cumulative incidence of breast tumor metastasis ranges between 21 to 32%. Despite extensive research on metastasis, our knowledge about the mechanism of metastasis is still elusive. Based on the “seed and soil” concept, developed by the English surgeon Stephen Paget over a century ago, selection of site for tumor cells to spread is not only determined by the tumor cells (seed), properties of the target organ (soil) also has huge impact (Fidler, 2003; Zhuyan et al., 2020). The microenvironment of target organs plays pivotal role in dissemination and colonization of tumor cells. Early and late dissemination patterns are two competitive models for metastasis. Based on the early dissemination model, metastatic cancer cells first seed an organ, forming tumors that stay dormant for a period of time. According to the late dissemination model, metastatic tumor cells arise as a result of Darwinian selection at later stage of tumors, during multistep carcinogenesis. These cancer cells are more invasive and mesenchyme-like. According to the late dissemination model, metastatic tumor cells first dissociate from their primary site and invade to around tissues, and then penetrate to the local vessels. After staying alive in circulation, they become lodged in the capillaries of target organs. They enter by extravasation and adapt to the novel

microenvironment in the new organ (Y. Feng et al., 2018; Riggi, Aguet, & Stamenkovic, 2018).

One of the defining characteristics of metastatic cancer cells is that they have traversed the epithelial-to mesenchymal transition (EMT), which is a reversible process in the metastatic trajectory of malignant cells (Lamouille, Xu, & Derynck, 2014). The EMT consists of a group of biological programs that are induced by alternative transcription factors. During the EMT, epithelial cells undergo morphological changes, including redirection from apical/basal cell polarity to a front/rear polarity, loss of epithelial properties, morphology changes, release lateral cell junctions and connections to the basal substrate, elongation, and acquisition of motile and invasive properties (Dudas, Ladanyi, Ingruber, Steinbichler, & Riechelmann, 2020). Tight junctions are protein complexes comprised of occludin, claudins, adhesion molecules, cingulin and zonula occludens (ZO). In epithelial tumors, TJ proteins preserve adhesiveness of cells in tumors mass and proliferation. Alterations in TJs can split a tumor mass, and initiate metastasis (Salvador, Burek, & Förster, 2016).

1.7.2 Role of drug resistance in breast cancer recurrence

Resistance to therapy is a leading challenge in oncology clinics, so understanding resistance should guide breast cancer treatment. ER and HER2 are two important prognostic factors for the treatment of breast cancer patients since resistance of patients to endocrine and HER2 targeted therapies can lead to disease relapse.

1.7.2.1 Endocrine resistance

Selective ER modulators (SERMs) such as tamoxifen, or selective ER downregulators (SERDs) like fulvestrant, and aromatase inhibitors (AIs) like letrozole, anastrozole, and exemestane, all have been used to treat ER positive breast cancer and reduce cancer recurrence and mortality (Wardell, Marks, & McDonnell, 2011). AIs and SERMs are being used to deplete estrogen levels in post-menopausal and pre-menopausal patients respectively; however a meta-analysis of the results from 88 trials involving 62,923 women with ER positive and early stage breast cancer, who had completed five years of endocrine therapy, show a persistent risk of recurrence and death from breast cancer for at least 20 years after the original diagnosis (Pan et al., 2017). Reactivation of the ER ligand leads to endocrine resistance that relies on gain-of-function mutations in ER, crosstalk between ER and growth factor receptors like EGFR or HER2 and activation

of oncogenic pathways like PI3K/mammalian target of rapamycin [TOR], RAS/RAF/MEK/ERK, and alteration of ER with its coactivators and corepressor (Hanker, Sudhan, & Arteaga, 2020).

1.7.2.2 HER2 targeted treatment resistance

Amplification or overexpression of HER2/ERBB2 in breast tumors is predictor of an aggressive phenotype and a poor prognosis. HER2 targeted drugs (i.e., monoclonal antibodies like trastuzumab and pertuzumab, antibody-drug conjugates like T-DM1, and tyrosine kinase inhibitors like lapatinib and neratinib) are being used to treat HER2 positive breast tumors and all of these agents can improve breast cancer patient survival (Yang, Li, Bhattacharya, & Zhang, 2019). A huge portion of HER2/ERBB2 amplified or overexpressed breast cancer patients can develop de-novo or acquired resistance to HER2-targeted therapies, including trastuzumab. Resistance leads to disease recurrence (Pohlmann, Mayer, & Mernaugh, 2009), e.g., almost 70% of patients acquire resistance to trastuzumab within 1 year of treatment. Proposed mechanisms for trastuzumab resistance have been proposed: (a) Complication of binding of trastuzumab to HER2 due to lack of extracellular domain and the binding site of trastuzumab in p95HER2 site of HER2 receptor. (b) Upregulation of HER2, downstream signaling pathways, (e.g., PI3K/AKT) due to loss of PTEN. (c) Signaling through alternate receptor pathways, like the Insulin-like growth factor-I receptor, Cdk inhibitor/p27 and EGFR, and HER3 (Pohlmann et al., 2009), and (d) activation of autophagy.

1.8 Role of autophagy in breast cancer

Autophagy is a survival mechanism for capturing, degradation and recycling of intracellular proteins and organelles in lysosomes (Doherty & Baehrecke, 2018), autophagy flux consists of few steps including; autophagosome formation, fusion of autophagosome with lysosomes, cargo degradation and release of macromolecules into the cytosol. Autophagy is linked to metabolic stress, genomic damage, and tumor formation (Singh et al., 2018). Cancer cells depend more on autophagy than normal cells, which may be because innate microenvironmental defects that are metabolically and biosynthetically demanding. Upregulation of basal autophagy in hypoxic tumor environments and RAS-transformed cancer cells was shown through numerous studies (Santana-Codina, Mancias, & Kimmelman, 2017; Yun & Lee, 2018). All aforementioned points suggest that autophagy promotes tumor development (White, 2015). Autophagy can inhibit tumor formation by diminishing oxidative

stress and promoting the DNA damage response and genome instability (Karantza-Wadsworth et al., 2007; Mathew et al., 2007). The role of autophagy in metastasis of all tumors, including breast tumors, is proven. Modulation of tumor cell motility and invasion, cancer stem cell differentiation, making the epithelial-to-mesenchymal transition, and dormancy of tumor cells through autophagy lead to metastasis and invasion of tumor cells (Mowers, Sharifi, & Macleod, 2017). Safely targeting of autophagy with autophagy inhibitors like chloroquine and hydroxychloroquine for cancer treatment has been tested in clinical trials (Levy, Towers, & Thorburn, 2017). Activation of autophagy was linked to trastuzumab and lapatinib resistance in breast cancer, and acted as a survival mechanism against HER2 targeted therapies with trastuzumab and lapatinib in breast cancer cells; and based on a recent preclinical and early clinical trials, combination treatment of cancer cells and patients with autophagy inhibitor chloroquine or hydroxychloroquine and HER2- targeted inhibitors may boost tumor cell death (Janser, Tschan, & Langer, 2019)

1.9 Breast cancer related data resources

Cancer research is in the age of “Big Data” (Clare & Shaw, 2016). The major purpose of precision medicine in cancer is providing the right drug for the right individual, based on their genetic background. For this purpose, genome-wide studies of large numbers of breast cancer patients were used to identify clinically significant variants and personalize treatment. Today use of “multi omics” analysis, in large, cancer cohorts have been a significant step forward (Low, Zembutsu, & Nakamura, 2018). The Cancer Genome Atlas (TCGA), The METABRIC (Molecular Taxonomy of Breast Cancer International Consortium), European Genome-phenome Archive are examples of using of omics technologies in a considerable number of normal breast tissues and tumors.

2 Aims

Breast cancer is a highly heterogeneous disease with frequent genetic rearrangements, such as copy number alterations (CNA), translocations, inversions, insertions, and deletions. Despite progress in targeted therapy tailored to the patient genetic background, disease can relapse 6 to 10 years after diagnosis. Finding prognostic and predictive markers in breast cancer is complicated because there are distinct molecular subtypes and a great degree of heterogeneity. Identifying novel genes that correlate with progression can lead to new drug targets for better outcome of patients. Inter-chromosomal or intrachromosomal changes of chromosome segments can generate fusion genes, some of which have oncogenic properties. Various oncogenic mechanisms have been described, such as oncogene activation, tumor suppressor deletion/downregulation, and the creation of novel proteins that can alter cellular pathways. Fusion genes can produce chimeric proteins or dysregulate expression of host genes and intragenic miRNAs. The aim of this study is to identify novel genes that support tumor progression in breast tissue and find defects in their sequence or activity for diagnosis, prognosis, and therapeutics. The method used to find new breast cancer genes was screening of fusion genes in large numbers of breast cancer cell lines and a large cohort of breast tumors.

2.1 Specific aims of the study

1. Use *in silico* analysis of fusion genes from breast tumors and breast cancer cell lines to identify novel breast cancer genes.
2. Explore the role of the identified genes in progression of breast tumors.
3. Analyze whether the roles of the identified genes support tumor progression in breast cancer cell lines.
4. Investigate how the pathways regulated by the identified genes can result in therapy-resistant cell lines.

3 Materials and methods

3.1 In silico analysis

A list of fusion genes from breast tumors were collected from three published papers (Asmann et al., 2012; S. Nik-Zainal et al., 2016; Yoshihara et al., 2015). In addition, a list of fusion genes from breast tumors from The Cancer Genome Atlas (TCGA) was purchased from MediSapiens (www.medisapiens.com). They used the MediSapiens Fusion SCOUT pipeline to identify fusion genes in RNA-Seq data from TCGA. Information about fusion genes in BC cell lines were collected from publications (Asmann et al., 2011; Edgren et al., 2011; Kalyana-Sundaram et al., 2012; Kangaspeska et al., 2012; Klijn et al., 2015; Robinson et al., 2011). Furthermore, we looked for fusion genes by analyzing RNA-Seq data of BC cell lines with the fusion finding algorithm SOAPfuse (Jia et al., 2013): CAMA-1 (GSM1172856), MDAMB134VI (GSM1172886), MDA-MB-231 (GSM1172889), SUM-225 (GSM1172901), SUM-229 (GSM1172902), SUM52 (GSM1172903), SUM44 (GSM1897347), and UACC893 (GSM1172907/GSM1897353). The paired-end RNA-Seq data from the cell lines were mapped to the human reference genome (hg19) and annotated transcripts (Ensembl release 75) using SOAP2. Then, SOAPfuse was used to identify fusion genes by detecting span and junction reads from the aligned data. Analyses of the RNA-Seq data from the cell lines also were purchased from MediSapiens. All identified fusion genes from tumors and cell lines were listed in Excel in two separate columns, and VLOOKUP function was used to retrieve common fusion genes in two columns. The following criteria were applied to common fusion genes among the 2 groups to merit further analyses: 1) have a similar breakpoint in breast tumors and cell lines, 2) be recurrent in tumors, 3) not be located within an amplicon carrying a known oncogene unless it was part of the fusion, and 4) possess a function supportive of tumorigenesis. Fusion genes that passed the criteria above were considered for validation.

3.2 Cell lines

To confirm fusion genes found in the *in-silico* analysis, the UACC893 (ATCC® CRL-1902™) cell line was purchased from American Type Culture Collection (ATCC) and MCF7 cell line was provided by the archive of Cell biology unit in the Department of Pathology of Landspitali University Hospital.

To investigate VMP's role in HER2 positive cell lines, BT474 (ATCC® HTB-20™) and MDA-MB-361(ATCC® HTB-27™) cells were also purchased from ATCC and compared to a control MCF10A cell line, provided by Dr Pórarinn Guðjónsson's laboratory. T47D, SUM52, and MDA-MB-231 cells also were provided from the Archive of Cell Biology in the Department of Pathology of Landspítali University Hospital. All cell lines were authenticated (genotyped) in our lab with related markers, to verify their identity and show they were free of contamination with other cells.

3.3 Cell culture conditions

UACC893 and MDA-MB-361 cells were cultured in Leibovitz's L-15 medium (Gibco™/ Catalog number: 11415-049) which were supplemented with 10% and 20% fetal bovine serum (Gibco™/ Catalog number: 101270-106) respectively. L-15 medium is used for growing of UACC893 and MDA-MB-361 cells, because its formulation was devised for use in a free gas exchange with atmospheric air. A CO₂ and air mixture are detrimental to cells when using this medium, so the cells were cultured in tissue culture flasks with a plug seal cap (Falcon™ / Corning, Catalog number: 353024). The T47D line was cultured in RPMI-1640 Medium (Gibco™/RPMI 1640 Medium, HEPES/Catalog number: 52400-025/ 500ml), supplemented with 10% FBS and 0.2 Units/ml bovine insulin (Sigma Aldrich/ Catalog number: I0516). MDA-MB-231 and MCF7 cells were cultured in RPMI1640 (Gibco™/RPMI 1640 Medium, HEPES/Catalog number: 52400-025/ 500ml) supplemented with 10% FBS. SUM52 cells were cultured in Ham's F12 (Gibco™/ Catalog number: 11765054), supplemented with 5% FBS and 2 mg/ml Insulin and 0.5 mg/ml Hydrocortisone. BT474 cells were cultured RPMI1640 medium supplemented with 10% FBS, 10 µg/ml insulin and Sodium Pyruvate (Gibco™/ Catalog number: 11360070). MCF10A cells were cultured DEMEM/F12 (Gibco™/ Catalog number: 11330-032), supplemented 5% horse serum (Gibco™/ Catalog number: 16050-122), 20 ng/ ml EGF, 0.5 mg/ml Hydrocortisone, 100 ng/ml Cholera toxin and µg/ml Insulin. Penicillin streptomycin (Gibco™/ Catalog number: 15140-122) was added to all cultures.

3.4 Breast tumors cohorts

Cohort 1 consisted of 158 BC patients, diagnosed between 1987 and 2003 (Appendix table 1); and cohort 2 consisted of 277 patients, diagnosed between 2003 and 2007 (Appendix table 2). The Nordic cohort consisted of 577 primary breast tumors from patients whose majority was diagnosed

between 1987 and 2003 in Finland, Sweden, and Iceland (including samples from cohort 1) (Jonsson et al., 2010; Reynisdottir et al., 2013). The TCGA cohort consisted of 818 BC patients, diagnosed between 1988 and 2013 (Ciriello et al., 2015) (Appendix table 3); and 2,509 METABRIC patients, diagnosed between 1980 and 2005 (Curtis et al., 2012; Pereira et al., 2016; Rueda et al., 2019) (Appendix table 4) with data available for both cohorts through cBioPortal (Cerami et al., 2012; J. Gao et al., 2013) and from Rueda et al. (Rueda et al., 2019). The METABRIC/EGA cohort is a subset of the METABRIC cohort that is included in the European Genome-Phenome Archive (EGA) for quantification of microRNAs. The cohort consists of 1286 patients, but our analysis subtracted patients with no histograde data (n=66) leaving us with the clinical information of 1220 patients for our analysis of hsa-mir-21-3p (Appendix table 5).

3.5 Clinical information and tumor characteristics

The relevant patient data for cohorts 1 and 2 were collected from hospital records at Landspítali, The National University Hospital of Iceland, as described previously (Gudmundsdottir et al., 2012). Primary fresh and frozen tumors were obtained from the Department of Pathology, as well as 35 non-neoplastic breast tissue samples, taken as far away from the tumor as possible. Informed consent was obtained from all patients involved in this study according to the national guidelines. The study was approved by The Icelandic Data Protection Commission (2001/523 and 2002/463) as well as the National Bioethics Committee of Iceland (99/051, 99/051_FS1, VSN-11-105, VSN-15-138). Survival data, including cause of death and time, was retrieved from patient records and the Icelandic National Register. Clinical pathology data, including age at diagnosis, tumor size, type, nodal status, ER, PR, HER2 and Ki67 levels were obtained from records at the Department of Pathology at Landspítali, The National University Hospital of Iceland. Immunohistochemistry staining was performed according to the manufacturer's protocol (Dako) and used to evaluate ER (M7047, clone 1D5, 1:100), PR (M3569, clone PR 636, 1:100), HER2 (HercepTest /K5207) and Ki67 (M7240, clone MIB1, 1:100) levels. Status of ER and PR was determined as being positive (staining 1+ to 3+) or negative (no staining). HER2 status was evaluated according to American Society of Clinical Oncology/College of American Pathologists guidelines, as being either positive (3+ staining plus 2+ staining and positive FISH analysis) or negative (1+ plus 2+ staining and negative FISH analysis) (Wolff et al., 2007). Ki67 data for tumors was scored, highly proliferative (number of cells and staining

intensity was above 10%) and slowly proliferative (number of cells and staining intensity was below 10%). Molecular classification of tumors was performed according to WHO histological classification (Hu et al., 2006); and the Bloom-Richardson system was used for histological grading of tumors.

3.6 DNA/RNA/miRNA extraction from tumors

Extraction of DNA and total RNA from breast tumor tissues in cohort 1 was performed with using TRIzol Reagent (Invitrogen/ Catalog number: 15596-018/ 200mL), extracting 10 to 100 mg. tissue per sample. All the tissues were stored in an -80 °C freezer and dissected inside a cooled Petri dish with cooled tweezers, on dry ice. Total RNA and DNA was extracted with TRIzol reagent, chloroform, isopropanol and 70% ethanol.

DNA and total RNA, from breast tumors and matched normal breast tissues in cohort 2, was extracted using the Qiagen AllPrep DNA/RNA/miRNA Universal Kit (Catalog number: 80224) from a maximum of 30 mg freshly frozen tissue. Tumor and normal breast tissue cells were disrupted and homogenized using the TissueLyser system. Tissues were added to lysis buffer (394 µL of RLT Plus lysis buffer, 4 µL of 2-Mercapthoethanol (14.3 M), and 2 µL Reagent DX) in a 1.5 ml Eppendorf tube containing 2 steel balls. The tubes were placed into the tissue lyser adapter (pre-cooled to -20°C) and tissues were sheared at 50 Hz, twice for 4 minutes (2x 4 min). The lysate was transferred to the QIAshredder column and spun at 14,800 rpm for 5 minutes at RT. The supernatant was applied to an AllPrep DNA column in a 2 ml Eppendorf tube, spun for 30 sec at 14,800 rpm, and eluate collected and stored at 4°C. For fatty tissues like breast, 90 µl of chloroform was added to the supernatant from previous step and spun at 4°C for 3 minutes at 14,800 rpm to separate to two phases. The water phase on top was transferred into a fresh 2 ml Eppendorf tube and 50 µl of Proteinase K was added to it, and mix by pipetting, then 200 µl of 96 to 100% ethanol was mixed well with the flow-through and left for 10 minutes incubation at room temperature, 700 µl of the sample, including any precipitate that may have formed was transferred to an RNeasy Mini spin column, placed in a 2 ml collection tube and centrifuged for 15 seconds at 14,800 rpm, the flow-through was discarded. After repeating this step, the entire sample was passed over a RNeasy mini spin column and washed with 500 µl Buffer RPE. After a 15 second spin on a microcentrifuge (at 14,800 rpm), the flow-through was discarded. After mixing 10 µl DNase I solution to 70 µl Buffer RDD the mix was added directly onto the RNeasy Mini spin column membrane and placed on the benchtop (20–30°C) for 15 min and

later 500 μ l Buffer FRN added to the RNeasy Mini spin column and spin for 15 sec at 14,800 rpm. The flow-through was saved and applied to the spin column and after centrifuging for 15 seconds at 14,800 rpm, the flow through was discharged. 500 μ l Buffer RPE was added to the RNeasy Mini spin column as a wash and centrifuged for 15 seconds at 14,800 rpm. Next, 500 μ l of 96–100% ethanol wash was applied, and the column centrifuged for 2 min at 14,800 rpm to remove the ethanol. This step was repeated. The RNeasy Mini spin column was then moved to a new 1.5 ml collection tube and RNA was eluted with 30–50 μ l of RNase-free water, added directly to the column membrane; after 5 minutes incubation at room temperature, the column was centrifuged for 1 min and the eluate collected. For genomic DNA purification, 350 μ l Buffer AW1 was added to the AllPrep DNA Mini spin column and centrifuged for 15 seconds at 14,800 rpm to wash the spin column membrane. The flow through was discarded and 60 μ l Buffer AW1 plus Proteinase K was added to the column membrane and incubated 5 min at room temperature. The column was washed with 350 μ l Buffer AW1 and centrifuged for 15 seconds at 14,800 rpm followed by 500 μ l Buffer AW2 and centrifugation for 2 min at 14,800 rpm. The AllPrep DNA Mini spin column was removed carefully from the collection tube and placed a new 1.5 ml collection tube. Finally, 100 μ l elution Buffer EB (preheated to 40°C) was applied directly to column membrane, incubated at room temperature (15–25°C) for 1 min, and collected by 1 min centrifugation at 10,000 RPM. The DNA concentration was measured by Nanodrop 1000 and the RNA quality was measured with Bioanalyzer 2100 RNA 6000 Nano kit (Agilent Technologies, cat. no. 5067–1511) as per the manufacturer's protocol. To assign integrity values for RNA, the RNA integrity number (RIN) algorithm was used. The majority of tumors had RIN \geq 8.

3.7 DNA/RNA/miRNA extraction from cell lines

Culture media was aspirated off the plated cells, and 1ml per 25cm² Tri Reagent (invitrogen/ Catalog number: 15596-018) was directly added. After a 5-minute incubation at room temperature, and pipetting up and down to homogenize, the lysate was transferred to 1.5 ml Eppendorf tube and 0.2ml per 1 ml Tri Reagent, chloroform was added, the sample was vortexed for 10–15 seconds, and incubated 10 minutes at room temperature. Tubes were then centrifuged at 13000 rpm for 18 minutes at 4°C, and the clear phase was collected and moved to new tube whereas the pink, lower phase was discarded. To the clear phase, we added 0.5ml of isopropanol (\geq 98%) per 1 ml TRIzol Reagent and, after inverting the tube a couple of times, the mixture

was incubated at room temperature for 10 minutes. The tubes were spun at 13000 RPM, for 12 min at 4°C, after which the liquid supernatant was discarded and the pellet dissolved in 1ml per 1ml TRIzol reagent, 70% ethanol. The mixture was vortexed, microfuged at 8000 RPM for 5 min, and the ethanol step was repeated. The pellet was collected, air dried briefly and dissolved in 10 to 30µl RNase-free water. The quantity of RNA was measured by Nanodrop 1000.

3.8 cDNA generation

For generation of cDNA the Revert Aid First Strand cDNA Synthesis Kit was used (Thermo Scientific/ Catalog number: K1622). On day one, 0.5µg of RNA aliquoted into a PCR plate and left at room temperature overnight to dry, on day two, 11µl of RNase free water was mixed with 1 µl of Random Hexamer Primer (100µM) and the 12 µl of mix was added on top of the samples in PCR plate and was put into a PCR thermal cycler. The following PCR procedure was performed:

65 °C	5 minutes
22 °C	10 minutes
4 °C	∞

cDNA reaction mix was prepared with adding 4 µl of 5x reaction buffer mixed to 1 µl of Ribolock RNase inhibitor (20U/µl), 2 µl of 10 mM dNTP mix and 1 µl of RevertAid M-MuLV RT (20020U/µl), 8 µl of mix was pipette into samples from previous stage and put on Thermal cycler. The following PCR procedure was performed.:

25 °C	5 minutes
42 °C	60 minutes
70 °C	5 minutes
4 °C	∞

3.9 RT-PCR

We performed RT-PCR with using 10ng/µl of cDNA generated from mRNA isolated from MCF7, SUM51, MDA-MB-231 and UACC893 cell lines to confirm expression of CCDC6: ANK3 and RPS6KB1:VMP1, GATAD2B: NUP210L, SMARCA4:CARM1, MYO6:SENP6, SUPT5H: SIPA1L3, ANKHD1: CYSTM and ITGB6: RBMS1 fusion genes from the *in-silico* analysis. For this purpose, a PCR mix was prepared according to the recipe in Table 1.

Table 1. PCR mix recipe for fusion genes confirmation

1. Betaine ($\geq 98\%$)	2.0 μl
2. Water	4.1 μl
3. Buffer (10x)	1.0 μl
4. dNTPs (100mM)	0.64 μl
5. MgCl_2 (25mM)	0.8 μl
6. Taq polymerase (5 U/ μl)	0.06 μl
7. Forward primer (20 pmol/ μl)	0.2 μl
8. Reverse primer (20 pmol/ μl)	0.2 μl

8.6 μl of PCR mix was added to 1 μl of cDNA template and the samples were subjected to the following protocol in a thermocycler:

- 94°C, 3 minutes
 - 94°C, 30 seconds
 - 60°C, 45 seconds
 - 72°C, 45 seconds
 - 72°C, 10 minutes
 - 4°C, ∞
- } x35 cycles

The sequence of forward and reverse primers used for confirmation of fusion genes with RT-PCR is shown in Table 2.

Table 2. Sequences of primers used for fusion-gene confirmation

Primer name	Sequence (5'-3')	T _m [°C]
CCDC6 -Fwd	TGCAGCAAGAGAACAAGGTG	60
ANK3-Rev	TGCTGACATTTCTTCCACGA	60
RPS6KB1-Fwd	GAAACTAGTGTGAACAGAGG	55.3
VMP1-Rev	CATAACTTTGTGCCATGGAG	55.3
GATAD2B-Fwd	AGATGATGTCCTGGCAAAGC	60
NUP210L-Rev	CCCCAGTTATGGTTGTTTGG	60
ITGB6-Fwd	GTTTCCTGCTCTCTGCAAGG	60
RBMS1-Rev	AGTGTCATTCCAGCCTCTCC	60
SMARCA4-Fwd	CTACCTCCACCCTCGGTGT	60
CARM1-Rev	GAACTGGATGAGGACGCTGT	60
MYO6-Fwd	GGATCTGTCCGAGCAGGAAG	56
SENP6-Rev	GGCTTGGCAGAAGAGTTTTG	56
SUPT5H-Fwd	TGTCAGCATTTCCAGTGAGC	56
SIPA1L3-Rev	ACCTTGCCTGTCAGATCCAG	56
ANKHD1-Fwd	TCTGCAACAGGAAACTGC	56
CYSTM-Rev	CCCATAGTGCTGAAGGTAGAGG	56

3.10 Gel electrophoresis

Samples were resolved on a 2% agarose gel in TBE buffer and visualized using ethidium bromide (Fluka BioChemika/GA12290). For loading, samples were mixed by adding 1µl 5x loading dye to 4µl PCR reaction, next to wells containing 1kb ladder and 50bp ladder molecular weight standards. Gels were run at 100V until the dye front was approximately 75 to 80% of the way down the gel. Images were collected using the Chemi XRS Documentation System from Bio-Rad, using UV-light. Images were used to estimate the quality of the PCR amplification product.

3.11 Sanger sequencing

To remove unwanted DNA and primers from the PCR product, exonuclease cleaning was performed prior to sequencing. A 5 µl cleaning master mix was prepared by adding 4.25 µl H₂O, 0.25µl Exol and 0.5µl FastAP to 2 µl of the PCR product. The PCR strips were then put inside a PCR thermal cycler programed for the following conditions:

Exonuclease cleaning program

- 37°C 15min
- 85°C 15min

A master mix for the sequencing reaction was prepared using BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems/ Catalogue number: 4336917). For each sample, two sequencing reactions were prepared, one with the forward PCR primer and the other with the reverse PCR primer. For making a master mix for the sequencing reaction, 2.5 µl H₂O, 1 µl 5xSeq.buffer and 0.5 µl BigDye (5x) were mixed together, 4 µl of sequencing master mix (5x), 1 µl of the cleaned PCR product. 0.1 µl (20 pmol/µl) of this mix was added to the PCR strips. The sequencing PCR reaction was run according to the below steps:

PCR sequencing reaction

- 96°C 10 seconds
 - 50°C 5 seconds
 - 60°C 4min
- } x35 cycles

Then 3.5 µl CleanSeq (MCLAB/CAT: BCB-100) and 20 µl 70% Ethanol were added to each PCR sequencing sample and the PCR strips were placed in a magnetic plate for 3-5 minutes. The ethanol mixture was discarded, and another ethanol mixture was pipetted to the mix and then also discarded. Finally, 50 µl 1xElution buffer was added to each sample, and after 3-5 minutes incubation, the PCR strips were replaced on magnetic plate for 3-5 minutes and the mixtures were pipetted into the sequencing plate. Sequencing runs were performed on a 3130xl Genetic Analyzer, Applied Biosystems, and data was analyzed with using the Sequencer program (Gene Codes Corporation).

3.12 Gene copy number variation analysis

VMP1 copy number variations analysis was carried out as described previously (Hoebeeck et al., 2005) in cohort 1 and cohort 2. We used Power up™ SYBR® Green Master Mix (Thermo Fisher /A25776) with VMP1-forward, and VMP1-reverse primers for amplification, and the ZNF80 and GPR15 genes used simultaneously as reference genes. The sequences of forward and reverse primers used for gene copy numbers variations analysis are shown in Table 3. VMP1 copy number data for Nordic (Jonsson et al.,

2012), TCGA (Cerami et al., 2012; J. Gao et al., 2013) and METABRIC (Pereira et al., 2016) were retrieved from GEO the dataset GSE22133 and cBioPortal respectively; and comparative genomic hybridization (CGH) on microarrays was used to measure copy number data. TCGA, and Affymetrix SNP6 arrays were used for measuring copy number data on METABRIC. Copy number variation (CNV) was defined according to methods described in the TCGA dataset (Network, 2012). VMP1 primers spanned the 3' UTR of VMP1 gene, which potentially detected MIR21 gene copy number data. For this reason, VMP1 CN data was used as MIR21 gene data.

Table 3. Primer sequences used for copy number variation analysis

Primer name	Sequence (5'-3')
VMP1-forward	GCACAAAGTTATGCCAAACG
VMP1-reverse	TCCCAATTTAAGGCAGAACC
GPR15-forward	GGTCCCTGGTGGCCTTAATT
GPR15-reverse	TTGCTGGTAATGGGCACACA
ZNF80-forward	CTGTGACCTGCAGCTCATCCT
ZNF80-reverse	TAAGTTCTCTGACGTTGACTGATGTG

For performing the VMP1 copy number analysis, 3 mixes were prepared by adding 5 μ l of PowerUp™ SYBR® Green Master Mix (2X) to 0.6 μ l of water, 2 μ l of Betaine, 0.2 μ l of forward and reverse primers of VMP1, GPR15 and ZNF80 genes. The mixes were pipetted to qPCR plates and 2 μ l of 0.5 ng DNA from sample was added. All samples were run in triplicate. The thermal cycling conditions for DNA amplification is listed in Table 4.

Table 4. Thermal cycle conditions for copy number analysis

Step	Temperature	Duration	Cycles
UDG Activation AmpliTaq Fast	50°C	2 minutes	Hold
DNA Polymerase, Up activation	95°C	2 minutes	Hold
Denature	95°C	15 seconds	40
Anneal/Extend	60°C	1 minute	

(Primer $T_m \geq 60^\circ\text{C}$)

The comparative CT method (also known as the $2^{-\Delta\Delta CT}$ method) was used to analyze the data, $2^{-\Delta\Delta CT} = \frac{[(CT \text{ gene of interest} - CT \text{ internal control}) \text{ sample A}]}{[(CT \text{ gene of interest} - CT \text{ internal control}) \text{ sample B}]}$, fold change = $2^{-\Delta\Delta CT}$. Based on measurements of DNA quantity in tumors, copy number categorizations were done according to the method used in TCGA ("Comprehensive molecular portraits of human breast tumours," 2012). According to their method, copy number values were thresholded as follows: Samples with a quantity of DNA ≤ -2.35 were defined as extra lost (≤ 1 copy), $\leq -0.54 - (-1.4) =$ loss (1-2 copies), $-0.54 - 1.3 =$ Neutral (2-3 copies), $\geq 1.3 - 2.2 =$ Gain (3-4 copies), $\geq 2.2 - 3.15 =$ amplification (4-6 copies) and $\geq 3.15 =$ high amplification (≥ 6 copies).

3.13 Quantitative reverse-transcription PCR (RT-qPCR)

We performed RT-qPCR using Taqman Gene Expression Master Mix (Thermo Fisher/ Catalog number: 4369016) and Taqman Gene Expression Assays spanning exons 10–11 (E10-11; Thermo Fisher Scientific, Taqman/Hs00978589_m1), a probe spanning exons 2 and 3 (E2-3; Taqman/Hs00978582_m1) and TaqMan® Gene Expression Assay/ERBB2/ Assay ID: Hs01001580_m1, (Thermo Fisher Scientific Catalog no: 4453320). TATA-binding protein (TBP, 1702071 Applied Biosystems) was used as a reference gene. For this purpose, a PCR mix was prepared (see Table 5).

Table 5. RT-qPCR recipe

1. Water	2.6 μ l
2. Gene Expression Master Mix	5 μ l
3. TBP (20X)	0.2 μ l
4. TaqmanGene Expression Assays	0.2 μ l

8 μ l of the above was mixed with 2 μ l of 0.5ng/ μ l cDNA. All reactions were done in triplicate using 42 cycles in Applied Biosystems Step One Plus Real-Time PCR system, the cycling conditions shown in Table 6.

Table 6. q RT-PCR cycling conditions

Steps	Temperature	Duration	Cycles
Holding			
1	50°C	2 minutes	1
2	95°C	10 minutes	1
Cycling			
1	95°C	15 seconds	40
2	60°C	1 minute	

VMP1 and ERBB2 expression was calculated relative to TBP: $2^{-(\text{mean Ct}_{\text{target}} - \text{mean Ct}_{\text{reference}})}$ and transformed with \log_2 . VMP1 mRNA in the Nordic data set; TCGA and METABRIC was measured with using gene expression microarrays.

3.14 cDNA generation for miRNA

cDNA synthesis for miRNA was performed using the cDNA Synthesis Kit II (Qiagen, Exiqon cat. no.203301), 2 μl of 5 ng/ μl RNA from tumor samples was mixed with 2 μl of 5x reaction buffer, 5 μl Nuclease free H_2O and 1 μl Enzyme mix and run in a Thermal Cycler program:

- 42°C 60:00
- 95°C 5:00
- 4°C ∞

The generated cDNA was diluted 80x with nuclease-free water before running the samples.

3.15 miRNA quantification with q-PCR

miRNA quantification was performed with using miRCURY LNA SYBR® Green PCR (QIAGEN /Cat No./ID: 339345/2x), and EXIQON/QIAGEN primer sets hsa-miR-21-5P (YP00204230) and hsa-miR-21-3p (YP00204302), hsa-miR-16-5P (YP00205702) were used as reference genes. The PCR mix was prepared according to the recipe shown in Table 7.

Table 7. PCR mix recipe for miRNA quantification

1.	ExiLENT SYBR green Master Mix2x reaction buffer	5µl
2.	mirCURYuniRT primer mix	1 µl
3.	Diluted cDNA with ROX	4µl

All reactions were performed in triplicate using 40 cycles of the thermal cycling program and a melting curve analysis (Table 8).

Table 8. q RT-PCR cycling conditions for miRNA quantification

Step	Temperature	Duration	Cycles
PCR initial heat activation	95°C	10 minutes	Hold
Denaturation	95°C	10 seconds	40
Combined annealing/extension	60°C	1 minute	
Melting curve analysis	60 -95°C		

Hsa-miR-21-3p and hsa-miR-21-5p expression was calculated relative to hsa-miR16-5p according to below formula.

$$\text{Normalized fold change} = \frac{2^{-(CT_{\text{target}} - CT_{\text{reference}})}}{2^{-(CT_{\text{target in positive control}} - CT_{\text{reference}})}}$$

Hsa-miR-21-3p and hsa-miR-21-5p in TCGA and European Genome-phenome Archive quantified were quantified with microarray technology.

3.16 Preparing Whole-Cell Lysates for Immunoblotting:

Preparation of whole-cell lysate from cells was carried out according to a protocol described by Larisa Litovchick (Litovchick, 2018). Growth medium was removed from confluent cells and they were rinsed two times with room temperature PBS and then 1ml of cold PBS with Halt™ Protease Inhibitor Cocktail (Thermo Scientific/ Catalog number: 87786) was added. The mixture was rocked over the cells and after removing cold PBS, 1 mL of SDS-PAGE sample buffer per 100-mm plate was added and plates were swirled to distribute buffer. The cells were collected using a scraper, and the extract was transferred to micro centrifuge tube on ice. Samples were heated 10 minutes at 95°C, chilled on ice, and centrifuged at 10,000g for 10 minutes; and the supernatant was transferred to new micro centrifuge tubes to use on western blot. SDS-PAGE Sample Buffer for Immunoblotting (6×) was prepared according to the recipe in table 9.

Table 9. SDS-PAGE sample buffer recipe

	Reagent	Quantity (for 10 mL)	Final concentration
1.2.1	Tris (2 M, pH 6.8)	1.8 mL	360 mM
2.2.1	SDS	1.2 g	12%
3.2.1	Glycerol	6 mL	60%
4.2.1	Bromophenol blue	1.5 mg	0.015%
5.2.1	β-mercaptoethanol (14.7 M)	1.8 mL	18%

3.17 Western Blot

For performing Western blot 18 µl of protein lysate was added to 2 µl of SDS-PAGE sample buffer and boiled at 95°C for 3-5 minutes, the mixture was spun briefly at 14000 RPM and loaded to 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (BIO-RAD/Catalogue number:4561085). Gels were run for 45 minutes at 170-200 volts and proteins transferred to Immobilon-FL PVDF, 0.45 µm membranes (Millipore/ IPFL00010) with using 20% transfer buffer and methanol. Membranes were blocked in Odyssey® Blocking Buffer (TBS) (Li-Cor/P/N927-60001) for 1 hour and subsequently with primary antibodies overnight at 4°C and the following day the membrane was washed with 4x TBS-T and incubated for 1 hour at room temperature with appropriate secondary antibodies, the images were captured using Odyssey CLxImager (LI-COR Biosciences) and analyzed using Image Studio 5.2. The antibodies used are listed in table 10.

Table 10. List of primary and secondary antibodies

Target	Manufacturer	Host	Dilution
VMP1	Cell signaling	Rabbit	1:500
HER2	Cell signaling	Rabbit	1:1000
ZO-1	Cell signaling	Rabbit	1:1000
E-CAD	BD Bioscience	Mouse	1:1000
ACTIN	Abcam	Mouse	1:1000
IRDye®680LT	LI-COR	Goat, anti-mouse	1:40,00 0
IRDye®800CW	LI-COR	Goat anti-Rabbit	1:25,00 0

3.18 Transfection with siRNA

Small-interfering RNAs (siRNA) targeting human VMP1 (Assay ID: s37756 / Catalog number: 4392420, Assay ID: s37755 / Catalog number: 4392420 and ERBB2 (Assay ID: s611 / Catalog number: 4390824 and Assay ID: s613/ Catalog number: 4390824) genes were purchased from Thermo Fisher. On day zero 60 to 80% confluent cells were seeded. The Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher/ Catalog number: 13778150) was diluted with Opti-MEM™ / Reduced Serum Medium (Gibco™/ Catalog number: 31985062) according to manufacture instruction, after adding Opti-MEM™ medium to 10µM siRNA, the diluted siRNA was mixed with diluted Lipofectamine™ RNAiMAX Transfection to a 1:1 ratio. After 5-15 minutes incubation, the siRNA-Lipid complex was added to the top of cells at the time of seeding. After 72-96 hours incubation in a 37 °C incubator the transfected cells were visualized with a microscope and q-RT-PCR and Western blot techniques were used to evaluate efficiency of knock down.

3.19 IncuCyte® Live Cell Analysis

Live cell analysis with IncuCyte also was used for assessment of the effect of silencing of VMP1, ERBB2 and VMP1+ERBB2 with siRNAs. A total of 3×10^4 of MDA-MB-361 and BT474 cells were seeded to a 96-well plate. siRNAs were added to the top of cells on the same day of seeding on 96-well plate. Twenty-four hours later, plates were moved to IncuCyte. Data regarding viability and the number of cells was collected every two hours and the cells were allowed to grow 144 hours. Media was changed every four days.

3.20 Apoptosis assay

To assess the effect of transfecting cells with siVMP1, siERBB2 and siVMP1+siERBB2, upon induction of apoptosis we used the Annexin V-FITC Apoptosis Detection Kit I (BD Pharmingen™/ CN: 556547). This kit is for quantitative determination of the percentage of cells undergoing apoptosis, using Annexin V-FITC and Propidium Iodide (PI). At 96 and 144 hours post transfection, cells were washed twice with cold PBS and then binding buffer added to a concentration of 1×10^6 cells/ml. After transfer of 100 μ l of the solution (1×10^5 cells) to a 5 ml culture tube, 5 μ l of Annexin V-FITC and PI was added. The mixture of cells, Annexin V-FITC, and PI was incubated at room temperature (25°C) in the dark, for 15 min. After adding 400 μ l of 1X binding buffer, samples were analyzed by flow cytometry. We had four siRNAs (siVMP1, siERBB2, and siVMP1 and siERBB2) and a scrambled siRNA was used as a negative control. All the treatments with siRNAs were performed in triplicate and results compared with Scramble siRNA. PI did not permeate to the membrane of viable cells with intact membranes and stained negative for it, whereas membranes of dead and damaged cells were permeable to PI. Apoptotic cells stained positive for both Annexin V-FITC and negative for PI. Necrotic cells stained positive for both Annexin V-FITC and PI. Live cells were negative for both Annexin V-FITC and PI. Cells in early apoptosis were Annexin V positive and PI negative while dead cells or in late apoptosis cells were both Annexin V and PI positive.

3.21 Statistical analysis

Statistical analyses were performed with R 4.0.3 (Chude & Amaravadi, 2017). Normalization of microarray DNA and mRNA measurements from the Nordic cohort, cohort 1, and cohort 2 with q-RT-PCR were performed by log 2 transformation of the data. miRNA measurements with q-RT-PCR in cohort 1 and cohort 2 were transformed as well with log2 and centered with subtraction of the median value from all measured values. All the DNA, mRNA, and miRNA measurements within TCGA, METABRIC, and METABRIC/EGA cohorts are already transformed in cBioPortal and EGA, with z-scores. All patient characteristics can be found in Tables in the Appendix (Appendix Tables 1 - 4) prepared by using the table function in R and percentages calculated for non-categorical clinicopathological variables. Normalized DNA, mRNA, and miRNA quantity correlation values were calculated using the Pearson's correlation coefficient test. Correlation analysis between mRNA and miRNAs levels with clinicopathological features of patients within study cohorts were performed with Student's t-test or

ANOVA. P-values below 0.05 were considered significant.

Survival analysis was performed with Kaplan-Meier and log rank test with the survminer package in R. All the hazard ratios were calculated with univariant and multivariant Cox regression analysis (Bradburn, Clark, Love, & Altman, 2003). For these analyses, VMP1 mRNA values in tumors from the Nordic cohort, cohort 1, cohort 2, TCGA and METABRIC cohorts were divided to high VMP1 (\geq mean + 1SD) and normal VMP1 ($<$ mean + 1 SD) groups. The rationality of this grouping was as follows: when VMP1 mRNA quantities were analyzed with histogram function in R, there was a group of patients with high VMP1 in histogram skewed to the right pulled away from the rest of patients. We postulated they might have different survival than rest of patients. hsa-miR-21-3p and hsa-miR-21-5p quantities in cohort 1, cohort 2, TCGA, and METABRIC/EGA cohorts were analyzed with the histogram function in R. All histograms of the four cohorts were according to Gaussian distribution. For this reason, tumors were divided into two high and low hsa-miR-21-3p groups, based on the median of hsa-miR-21-3p levels.

4 Results

4.1 Identification of potential breast cancer genes (Paper 1)

Chromosomal rearrangements like fusion genes represent a frequent genetic mechanism for oncogene activation in cancers (Anca Botezatu, 2016). Fusion genes have been shown to have role in tumor development, and can consequently result in inappropriate expression of gene partners involved in the fusion (Zimmerman et al., 2017). To identify potential novel breast cancer genes associated with progression, fusion genes in both breast cancer cell lines and tumors were compared.

4.1.1 Common fusion genes within breast cancer cell lines and tumors

Fusion genes are not unique features of cancer and exist in normal cells as well (Babiceanu et al., 2016). Breast cancer cell lines are free of the confounding effects of contamination by normal epithelial or nonepithelial cells (Neve et al., 2006) and allowed us to build a model to compare fusion genes in both breast cancer cell lines and tumors.

A list of gene fusions identified in breast cancer cell lines was retrieved from the published literature and added to a list gene fusion identified via the SOAPfuse and Medisapiens fusion-finding algorithms. One hundred eighty-three fusion genes (Appendix table 6) from 45 breast cancer cell lines (Appendix table 7) were compared to 5319 fusion genes identified in 1724 breast tumors (Asmann et al., 2012; Nik-Zainal, 2016; Yoshihara et al., 2015). Most tumors were from The Cancer Genome Atlas (TCGA); and none of the fusion genes in tumors were specific for any particular subtype. Fusion genes in both cell lines and tumors were identified from RNA-Seq data analyzed with the fusion-finding algorithms, SOAPfuse and MediSapiens (www.medisapiens.com). The majority of the 45 cell lines were ER⁻/HER2⁻ (n=17/45, 37.77%), followed by ER⁺/HER2⁺ (n=6/45, 13.33%), ER⁻/HER2⁺ (n=12/45, 26.66%), and ER⁺/HER2⁻ (n=8/45, 17.77%). The status of ER and HER2 for 2 of the cell lines was unknown. The number of fusion genes in cell lines was 4 (8.2%) per cell line and 3.08 (0.3%) per tumor sample. The MCF7 cell line had more fusion genes per cell line compared to others. Breast cancer cell lines and breast tumors had 15 fusion genes in common (Figure 7).

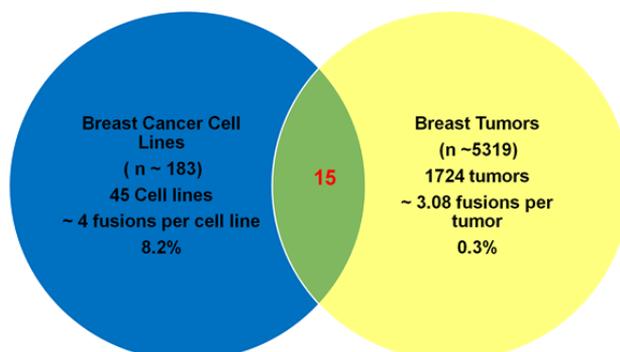


Figure 7. Breast cancer cell lines and tumors had 15 fusion genes in common. Fusion genes (n=183) from breast cancer cell lines (n=45) retrieved from publications and result of SOAPfuse and Medisapiens fusion finding algorithms. The fusion genes (n=5319) from breast tumors (n=1724) collected from publications and TCGA data analyzed with SOAPfuse and Medisapiens algorithms.

4.1.2 Verification fusion genes in common between breast tumors and breast cancer cells

Reverse transcription–polymerase chain reaction (RT-PCR) was used for verification of 15 fusion genes in common among breast cancer cell lines and tumors. We verified RPS6KB1:VMP1, CCDC6:ANK3, ITGB6:RBMS1, SMARCA4:CARM1, GATAD2B:NUP210L (Appendix figure 1) and MYO6:SENP6, SUPT5H:SIPA1L3 and ANKHD1: CYSTM fusion genes (Appendix figure 2; Table 11). There was difficulty with verification of VGLL4:SH3BP5 and ESR1:CCDC170 fusion genes; and we got multiple bands in gel electrophoresis. The other fusions had been verified to exist by others. The MYH9:EIF3D fusion gene was verified by Asman et al. (Asmann et al., 2011). ESR1:CCDC170 was verified by Wang XS et al. (Veeraraghavan et al., 2014). PLXND1:TMCC1 and INTS4:GAB2 were verified by Stephens et al. (Stephens et al., 2009). TIAM1:NRIP1 was verified by Schulte et al. (Schulte et al., 2012). POLA: CAPN1 was verified by Robinson et al. (Robinson et al., 2011). The precise PCR banding pattern was used to confirm the fusion gene size via gel electrophoresis.

Table 11. Verified 15 common fusion with RT- PCR

Gene A	Gene B	Breast Cancer Cell Lines
ESR1	CCDC170	ZR751
RPS6KB1	VMP1	MCF7
CCDC6	ANK3	UACC893
PLXND1	TMCC1	HCC1187
ITGB6	RBMS1	UACC893
MYH9	EIF3D	MCF7
VGLL4	SH3BP5	T47D
SMARCA4	CARM1	MCF7
GATAD2B	NUP210L	MCF7
MYO6	SENP6	MCF7
SUPT5H	SIPA1L3	SUM52
ANKHD1	CYSTM	MDAMB231
TIAM1	NRIP1	ZR7530
INTS4	GAB2	HCC2157
POLA2	CAPN1	HCC1806

4.1.3 Ten genes identified as putative breast cancer genes

To choose a gene candidate, the thirty genes that constitute the 15 fusion genes were filtered according to these criteria: a) an identical breakpoint in breast tumors and cell lines, b) recurrent in tumors, c) not located within an amplicon carrying a known oncogene unless it was part of the fusion and d) possessing a function supportive of tumorigenesis (available through publications). After applying these criteria, five pairs of fusion genes or 10 single genes remained. The frequency of their occurrence and the cell lines in which they were identified are listed in Table 12. ESR1:CCDC170 (n=11) was the most recurrent fusion gene within breast tumor samples (n=1724), followed by RPS6KB1:VMP1 (n=5), CCDC6:ANK3 (n=2). Although GATAD2B:NUP210L (n=1) and ITGB6:RBMS1 (n=1) fusion genes were not recurrent in breast tumors they were recurrent in other tumors. Two out of 11 ESR1:CCDC170 fusions were out of frame for both ESR1 and CCDC170 genes and nine were in the 5' UTR-CDS (coding regions). Three out of five RPS6KB1:VMP1 fusions were out of frame and two in-frame; the CCDC6:ANK3 fusion appeared once in-frame and once in the CDS-3' UTR; GATAD2B:NUP210L and ITGB6:RBMS1 fusions were in 5' UTR-CDS and in-frame, respectively. The total number of fusion genes in both breast tumors and cell lines, after filtering, is shown in table 12.

Table 12. Five fusion genes passed filtering criteria *

5' fusion gene partner	3' fusion gene partner	Number of fusions in tumors (%) ^a	Cell lines ^b
CCDC6	ANK3	2 (0.12%)	UACC893
ESR1	CCDC170	11 (0.64%)	ZR751
GATAD2B	NUP210L	1 (0.06%)	MCF-7
ITGB6	RBMS1	1 (0.06%)	UACC893
RPS6KB1	VMP1	5 (0.29%)	MCF-7

* The filtering criteria were the following: a) possessing identical breakpoint in breast tumors and cell lines, b) being recurrent in tumors, c) not located within an amplicon carrying a known oncogene unless it was part of the fusion and d) possessing a function supportive of tumorigenesis (available through publications). ^a the total number of tumors was 1724. ^b the total number of cell lines was 45.

The five fusion genes in common between tumors and cell lines were also analyzed with respect to other tumor types through (www.tumorfusions.org/2014 release); and the GATAD2B:NUP210L fusion gene was found once in lung adenocarcinoma (LUAD) and ovarian tumors (OV) in 5' UTR-CDS (coding regions). ITGB6:RBMS1 was found in three bladder cancer (BLCA) tumors as an in-frame fusion. CCDC6:ANK3 was found in one lung adenocarcinoma (LUAD) in 5' UTR-CDS (coding regions) and four ovarian tumors (OV), for which one was out of frame, two were in-frame and one was in the 5' UTR-CDS (coding regions), RPS6KB1:VMP1 was found in one bladder cancer (BLCA) tumor as an in-frame fusion gene, three head and neck squamous cell carcinoma (HNSC) which all were out of frame, eight lung adenocarcinoma (LUAD) tumors of which four were out of frame and four in-frame and one ovarian tumors (OV) as in-frame fusion.

4.1.4 Junction site of fusion genes verified

For verification of existence of predicted sequence of five fusion genes through fusion finding algorithms, the resulting PCR products of CCDC6:ANK3, GATAD2B:CCDC170, ITGB6:RBMS1 and RPS6KB1:VMP1 fusion genes were sequenced to verify the fusion junction. For each fusion pair, the forward primer of gene A and the reverse primer of gene B was used to amplify the sequence of fusion gene junction site. The validated junctions of the gene fusions are shown in Figure 8. We had difficulty in verification of the junction of the ESR1:CCDC170 fusion gene but it had been validated by sequencing by Wang XS et al. (Veeraraghavan et al., 2014).

comparative genomic hybridization (CGH) on microarrays and mRNA was measured with gene expression microarrays. The Nordic and TCGA cohorts consist of 577 and 818 patients, respectively. Based on Pearson correlation test results for the quantity of DNA and mRNA of 10 genes, VMP1 ($r=0.70$), RPS6KB1 ($r = 0.83$), GATAD2B ($r = 0.54$) and CCDC6 ($r = 0.66$) had the highest r values in TCGA cohort. CCDC6 was not amplified in either Nordic or TCGA cohorts and GATAD2B was only amplified in the TCGA cohort but not in the Nordic cohort. Genes amplified in both cohorts might have stronger oncogenic properties. The VMP1 and RPS6KB1 genes were the most frequently amplified genes within the two cohorts (Table 13).

Table 13. Amplification and correlation between DNA and mRNA of the gene partners that constitute the five fusion genes

Gene	**Amplification (%)		Correlation (r)	
	Nordic	TCGA	Nordic	TCGA
ANK3	0.29	2.8	0.20 [*]	0.39 [*]
CCDC170	0.6	1.9	0.07	0.03
CCDC6	0.9	1.5	0.52 [*]	0.66 [*]
ESR1	0.29	2.5	0.04	-0.05
GATAD2B	3.07	12.7	0.55 [*]	0.54 [*]
ITGB6	0	0.9	-0.01	0.15 [*]
NUP210L	0.92	12.9	0.01	0.16 [*]
RBMS1	0.93	0.6	NA	0.31 [*]
RPS6KB1	6.97	10.9	NA	0.83 [*]
VMP1	7.4	10.7	0.45 [*]	0.70 [*]

In both cohorts, DNA and mRNA were quantified by comparative genomic hybridization (CGH) and gene expression microarrays, respectively. The number of tumors with measurements for both DNA and mRNA in Nordic tumors is ($n = 337$) and in TCGA is ($n = 421$). *Denotes a significant result, $p < 0.05$. **Amplification refers to > 4 copies of genes.

4.1.6 Vacuole membrane protein 1 chosen as a potential breast cancer gene for further studies

RPS6KB1 has been implicated as a BC gene through studies in breast cancer patients and breast cancer cell lines (Holz, 2012; Noh et al., 2008; Pérez-Tenorio et al., 2011; van der Hage et al., 2004). For this reason, we explored the role of VMP1 in the survival of breast cancer patients in two cohorts, TCGA and Nordic. To compare the overall survival data among the Nordic and TCGA cohorts, patients were sorted into two groups based on their VMP1 mRNA levels: high expressors (\geq mean + 1 SD) and normal

expressors ($< \text{mean} + 1 \text{ SD}$) group. In the TCGA cohort, the median overall survival of patients expressing high and normal levels of VMP1 mRNA was 11.68 ± 4 years and 6.62 ± 2.49 years, respectively. In the Nordic cohort, the median overall survival of patients with high VMP1 mRNA was 12.6 ± 5.56 years, and 16.3 ± 5.18 years in patients expressing a normal amount of VMP1 mRNA. In TCGA and Nordic cohorts, the hazard ratio and confidence interval were $\text{HR} = 2.10$, $\text{CI} (1.09-4.04)$ and $\text{HR} = 1.37$, $\text{CI} (0.98-1.91)$, respectively. Log rank p-values are included in Figure 3. That the two cohort's survival curves are differently shaped may reflect the different numbers of death in each cohort. Survival data regarding the VMP1 gene made it strong candidate for follow up studies. Taken together, the analysis and results described above suggest that VMP1 is one of the strongest candidates to come out of the screen and worthy of following up in further studies.

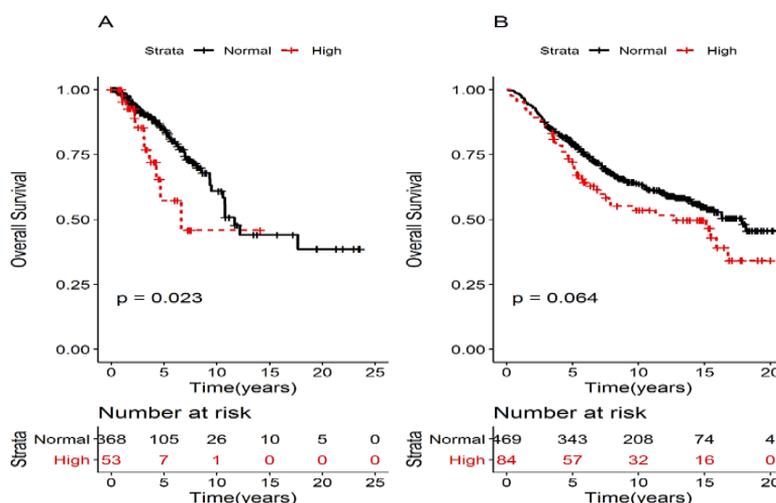


Figure 9. High VMP1 was marker of shorter overall survival.

Based on VMP1 mRNA levels in the patients' tumors, they were divided into two groups: tumors expressing high levels of VMP1 ($\text{high} \geq \text{mean} + 1 \text{ SD}$) and normal levels of VMP1 ($\text{normal} < \text{mean} + 1 \text{ SD}$). Overall survival (OS) of patients in (A) TCGA and (B) the Nordic cohort was examined with respect to quantity of VMP1 mRNA. The hazard ratio and confidence interval for TCGA cohort was ($\text{HR} = 2.10$, $\text{CI} (1.09-4.04)$) and for Nordic cohort was ($\text{HR} = 1.37$, $\text{CI} (0.98-1.91)$).

4.2 High expression of the vacuole membrane protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer (Paper 1)

As pointed out in the aims of this study, due to heterogeneity of breast cancer and relapse of disease 6-10 years after diagnosis, it is crucial to find novel breast genes associated with progression of disease that may be used as

diagnostic markers and as drug targets, like HER2. Due to role of fusion genes in oncogene activation and dysregulation of genes form them, the first part of our fusion-gene study in breast cancer cell lines and large tumor data sets were screened, identifying a single gene, VMP1. This paper represents data how was VMP1 found through screening of fusion genes and exploration of the role of VMP1 expression in two Icelandic breast tumor cohorts and following the results in METABRIC, TCGA cohorts.

4.2.1 VMP1 mRNA levels were high in breast tumors

In breast cancer tissue-based studies, non-cancerous breast tissue is often used as control for comparison for gene or protein expression, in both cancerous and non-cancerous tissues. During the process of breast cancer surgery, non-cancerous breast tissue usually is a benign histological tissue adjacent to the tumor that is resected at the time of surgery to remove cancerous tissue. This non-cancerous breast tissue (called *normal* in breast cancer tissue studies) includes unaffected, benign breast cells. To compare VMP1 mRNA expression within normal and breast tumors, we used 35 normal breast tissues, adjacent to matched tumors from cohort 2. VMP1 mRNA expression values in tumors were significantly higher than matched normal tissues (P value= $4.047e^{-10}$). Data was followed in TCGA, using data from 174 tumors and matched normal tissues. TCGA data was supportive of our data (P value= $1.5e^{-10}$; Figure 10).

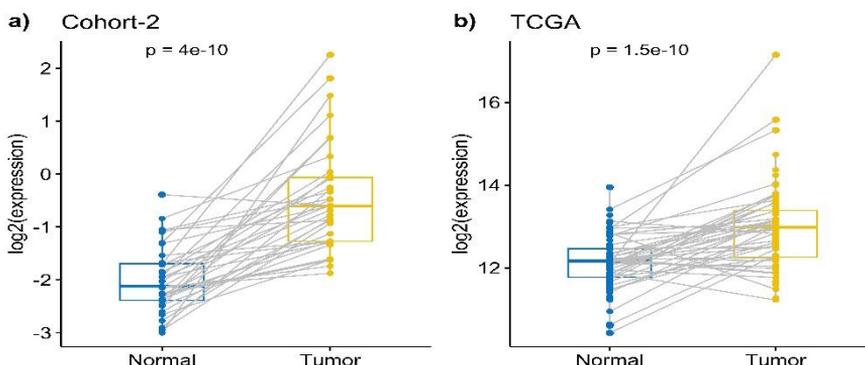


Figure 10. VMP1 mRNA levels were higher in breast tumors than paired normal tissues.

A. VMP1 mRNA levels were compared in breast tumors ($n=35$) and matched, normal breast tissues ($n=35$), in cohort 2. VMP1 mRNA data were normalized with log 2 transformation. Paired t-test in R was used to compare the normal and tumor groups. The p value was $4.047e^{-10}$. B. VMP1 mRNA levels were compared within tumors ($n=174$) and matched normal breast tissues from the TCGA, the p value was $1.5e^{-10}$.

4.2.2 VMP1 mRNA levels were higher in VMP1 amplified tumors

The impact of copy number variation (CNV) on protooncogenes expression mostly led to tumor development (Shao et al., 2019). Amplified genes are not always accompanied by elevated gene expressions and overexpression of a gene in amplified tumors makes them strong candidate for being driver (Ohshima et al., 2017).

Since VMP1 mRNA is associated with VMP1 CNVs within cohorts, the quantity of VMP1 DNA in cohorts 1 and cohort 2 was measured by qPCR. VMP1 DNA data for TCGA and METABRIC cohorts were retrieved through cBioPortal. Comparative genomic hybridization (CGH) on microarrays was used to measure VMP1 DNA in both TCGA and METABRIC cohorts (Pereira et al., 2016). In all of four cohorts, CNV was defined as amplification, gain, neutral and loss based on method used in TCGA data set ("Comprehensive molecular portraits of human breast tumours," 2012). According to this method amplification refers to 4-6 copies of DNA, gain 3-4 copies of DNA, neutral 2-3 copies and loss ≤ 1 copy of DNA. The frequency of VMP1 copy number variations was similar in cohorts 1 and 2. The frequency of VMP1 copy number variations in cohorts 1 and 2, TCGA and METABRIC are shown in Table 14. VMP1 mRNA was associated with VMP1 copy number variation in cohort1 ($P= 3.23e^{-10}$), cohort2 ($P =1.22e^{-13}$), TCGA ($P <2e^{-16}$) and METABRIC ($P <2e^{-16}$) (Figure 11). Elevated expression of VMP1 within study cohorts in VMP1 amplified tumors makes it strong candidate to be a cancer driver gene.

Table 14. VMP1 CNV frequencies within cohorts

Cohorts	Amplification	Gain	Neutral	Loss
Cohort1 (n=143)	4.89% (n=7)	9.09% (n=13)	72.72% (n=104)	13.28% (n= 19)
Cohort2 (n=273)	4.21% (n=12)	9.47% (n=27)	77.89% (n=222)	8.42% (n= 24)
TCGA (n=816)	10.78% (n=88)	29.77% (n=243)	45.34% (n=370)	14.09% (n=115)
METABRIC (n=1980)	10.45% (n= 207)	14.31% (n= 286)	63.78% (n=1263)	11.31% (n=224)

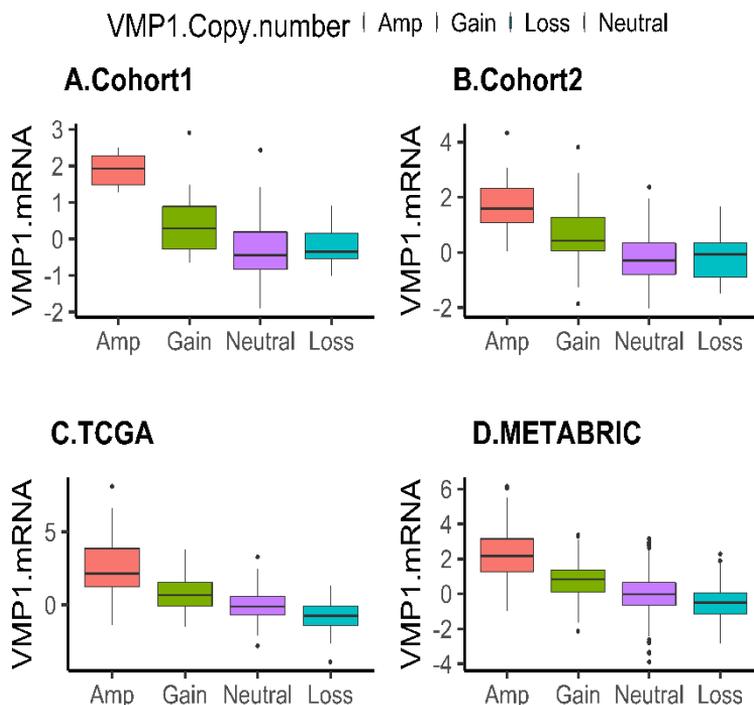


Figure 11. VMP1 mRNA associated with VMP1 copy number variation. VMP1's DNA levels from A) Cohort 1, B) Cohort 2, C) TCGA and D) METABRIC were categorized as Amplification, Gain, Neutral, and Loss (explained in detail in materials and method section). VMP1 mRNA was compared with ANOVA in R with copy number variations. The p values for cohort 1 was (P value= $3.23e^{-10}$), cohort 2 (P value= $1.22e^{-13}$), TCGA (P value $<2e^{-16}$) and METABRIC (P value $< 2e^{-16}$). The numbers and frequencies of VMP1 CNVs in cohort 1 (n=143) was 4.89% amplification (n=7), 9.09% gain (n=13), 72.72% neutral (n=104) and 13.28% loss (n= 19). In cohort 2 (n=274), the profile showed 4.21% amplification (n=12), 9.47% gain (n=27), 77.89% neutral (n=222) and 8.42% loss (n= 24). In TCGA (n=816), was: 10.78% amplification (n=88), 29.77% gain (n=243), 45.34% neutral (n= 370) and 14.09% loss (n=115). In METABRIC (n=1980) was: 10.45% amplification (n= 207), 14.31% gain (n= 286), 63.78% neutral (n=1263) and 11.31% loss (n=224).

4.2.3 Co-amplification of VMP1 and ERBB2 genes

The chromosomal 17q23.1 locus, where VMP1 resides, is amplified in 20% of ERBB2-amplified tumors (Haverty et al., 2008; Jonsson et al., 2010; Staaf et al., 2010). For this reason, the CNV of VMP1 DNA was analyzed with respect to ERBB2 CNV, within study cohorts. In cohort 1 (n=163), only three (14.28%) out of 21 tumors ERBB2 with amplified also had VMP1 amplification (P value = 0.08). In TCGA (n=817), VMP1 was amplified in 43 of

106 (40.5%) ERBB2-amplified tumors (n=106; $P < 2.2e^{-16}$). In METABRIC (n=1980), 31,2% (n=93) of ERBB2 amplified tumors (n= 298) had VMP1 amplification ($P < 2.2e^{-16}$). Cohort 2 did not have measurements for ERBB2 DNA. Based on these data 14.24%-40.56% of ERBB2-amplified tumors had VMP1 amplification, which confirms co-amplification of VMP1 and ERBB2 loci.

4.2.4 VMP1 mRNA was higher in HER2 positive tumors

VMP1 mRNA was associated with the HER2 status of tumors (as determined with immunohistochemistry) in cohort 1 ($P= 7 \times 10^{-4}$), cohort 2 ($P=0.004$), TCGA ($P= 0.003$) and METABRIC ($P < 2 \times 10^{-16}$) cohorts and was expressed more in HER2 positive tumors than negative tumors (Figure 12). Based on molecular subtype classification data, which classifies breast tumors to five subtypes (Luminal A, Luminal B, HER2 enriched Basal and Normal like), using immunohistochemistry data to measure expression of ER, PR, HER2, and Ki67. VMP1 mRNA was expressed at the highest levels in HER2 enriched and Luminal B subtypes and lowest levels in Basal subtype in cohort 1 ($p= 5 \times 10^{-6}$) TCGA ($p=2 \times 10^{-12}$) and METABRIC ($p < 2 \times 10^{-16}$) (Figure 13). HER2-enriched and Luminal B subtypes are the most common and Basal the rarest, among HER2-positive tumor subtypes. There is no information regarding molecular subtypes in cohort 2.

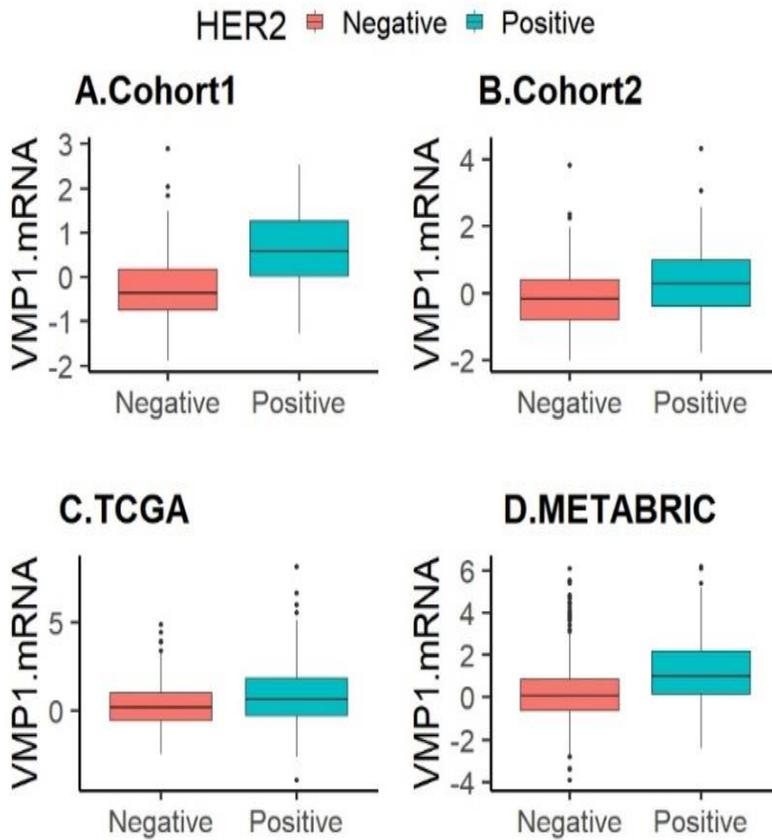


Figure 12. VMP1 mRNA expression associated with HER2 expression. VMP1 mRNA levels within A) Cohort 1, B) Cohort 2, C) TCGA and D) METABRIC were compared with HER2 status of tumors as determined by immunohistochemistry. Student t-test in R was used to compare VMP1 mRNA in HER2 positive and HER2 negative tumors. Based on t-test results p values within cohorts were, cohort 1(P value= 7×10^{-4}), cohort 2(P value= 0.004), TCGA (P value= 0.003) and METABRIC (P value $< 2 \times 10^{-16}$).

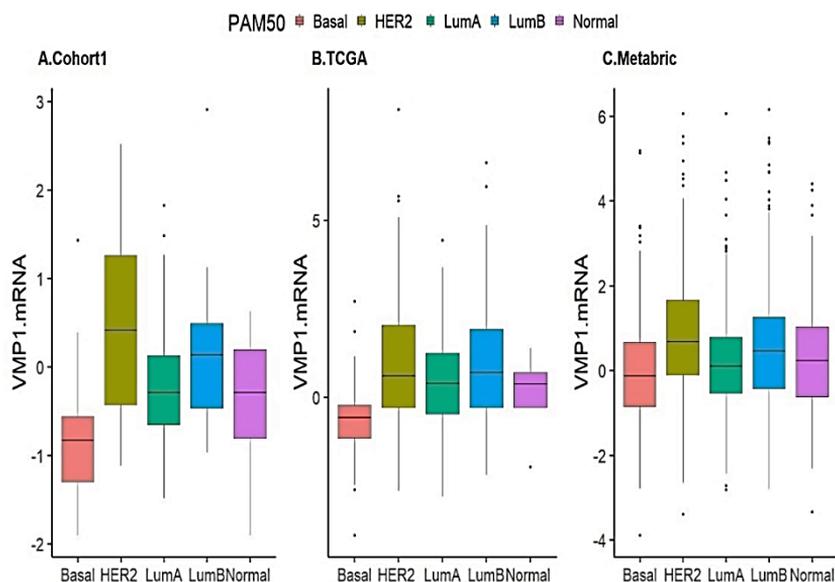


Figure 13. VMP1 mRNA levels are highest in HER2-enriched and Luminal B subtypes.

VMP1 mRNA levels within A) Cohort 1, B) TCGA and C) METABRIC were compared within molecular subtype status of tumors categorized (Basal=orange, HER2 enriched= evening green, Lum A = Green, Lum B = Blue, Normal like = Purple) with Hu et al.(Hu et al., 2006) method in cohort1 and PAM50 method in TCGA and METABRIC. VMP1 mRNA was compared with ANOVA in R with molecular subtype. Cohort 1 (P Value= 5×10^{-6}), TCGA (P Value= 2×10^{-12}) and METABRIC (P Value $< 2 \times 10^{-16}$).

Association of VMP1 mRNA with other clinic pathological factors within study cohorts is shown in the Appendix (Tables 8, 9, 10, and 11). In cohort 1, the quantity of VMP1 mRNA was higher in tumors that metastasized than tumors that had not ($p = 0.03$). In Cohort 2, the VMP1 mRNA quantity was high in grade 3 tumors ($p = 0.04$). In TCGA, it was high in ER positive ($p = 7 \times 10^{-6}$) and PR positive tumors ($p = 0.008$) and node-negative tumors ($p = 0.01$). In METABRIC, VMP1 mRNA was high in ER-positive tumors ($p = 0.01$), and IDC tumors ($p = 0.02$). In all cohorts, VMP1 mRNA was significantly higher in HER2-positive tumors compared to HER2-negative and was high in the ERBB2 subtype.

4.2.5 High VMP1 mRNA expression is associated with poor survival

Survival analysis for VMP1 mRNA status used breast-cancer-specific survival (BCSS) rather than overall survival (OS) because OS may reflect death due

to other diseases in addition to BC. Cohort 1, cohort 2, TCGA and METABRIC cohorts were divided into two according to VMP1 mRNA expression. High expression was based on the mean, plus one SD and all other samples were in the normal group. Median time of BCSS of patients with high and normal VMP1 mRNA were 13.22 ± 5.01 years and 3.75 ± 4.62 years in cohort 1. The hazard ratio was 2.31 and confidence interval was (1.27–4.18). In METABRIC cohort median time of BCSS in VMP1 high and normal groups was 23.5 years and 21.7 years, respectively. The hazard ratio and confidence interval in METABRIC cohort were 1.26, (1.02–1.57), respectively. The log rank p-value in cohort 1 was 0.0045 and in METABRIC was 0.032 (Figure 14). VMP1 mRNA did not associate with BCSS within cohort 2 (log rank $p=0.49$) and TCGA (log rank $p=0.12$). HER2 is a strong oncogene and VMP1 expression was high in HER2-positive tumors. For this reason, the effect of HER2 expression the VMP1 mRNA level was checked with Cox regression analysis in the cohorts 1 and METABRIC. In cohort 1 BCSS survival remained significant after adjusting for HER2 (HR: 2.03, CI (1.00–3.72)) but in METABRIC the effect of high levels of VMP1 mRNA was gone after adjusting for HER2 (HR: 1.03, CI (0.82–1.30)).

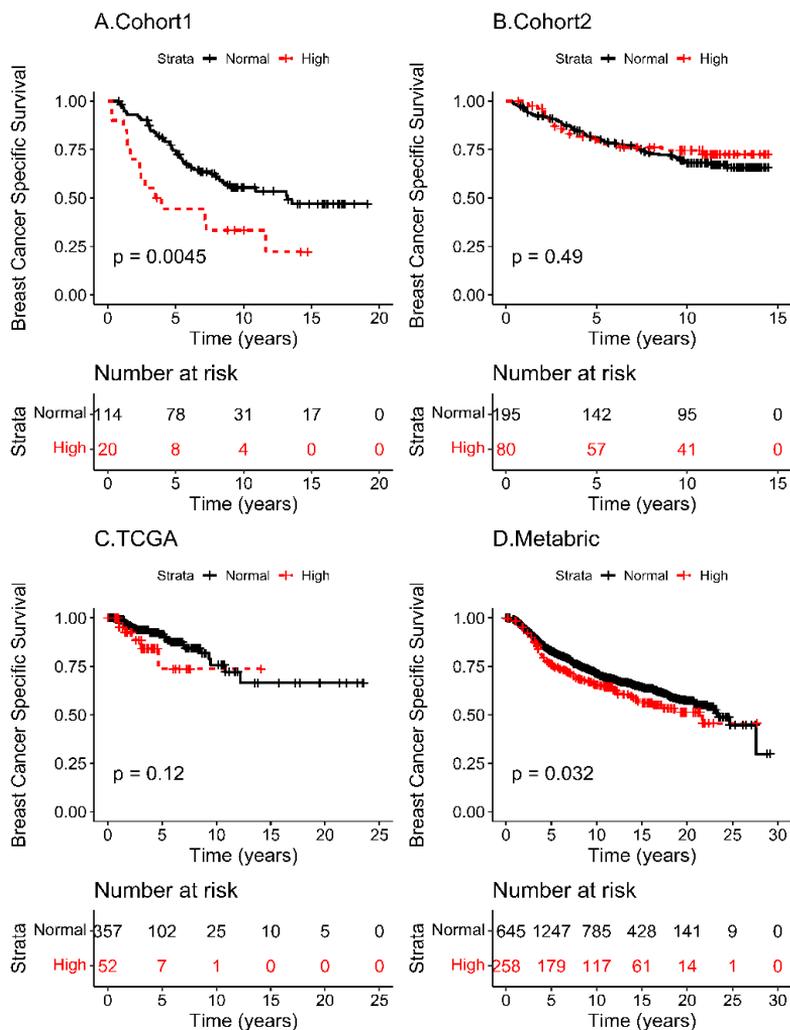


Figure 14. High expression of VMP1 was associated with shorter Breast Cancer Specific Survival (BCSS) in cohort 1 and METABRIC.

VMP1 mRNA levels within cohort 1, cohort 2, TCGA and METABRIC were divided into two groups, VMP1 high (\geq mean + 1 SD) and VMP1 normal ($<$ mean + 1 SD). The log rank p values and number of patients at risk are shown in the graphs. The BCSS hazard ratio (HR) in cohort 1 was 2.31 (CI 1.27–4.18), and after adjusting for HER2 expression the HR was 2.03 (CI 1.00–3.72). In METABRIC HR was 1.26 (CI 1.02–1.57) and after adjusting for HER2 expression it was HR = 1.03 (CI 0.82–1.30).

Due to the role of VMP1 plays in the initiation of autophagy (Molejon, Ropolo, & Vaccaro, 2013) and the high degree of autophagy in metastatic tumors (Galluzzi et al., 2015), VMP1 mRNA was analyzed with respect to distance recurrence free survival (DRFS) in cohort1 and METABRIC (Figure

15) for which there were data. In cohort 1 and METABRIC, elevated levels of VMP1 mRNA were associated with shorter time of distance recurrence. The hazard ratio (HR) and log rank p value and CI for DRFS in cohort 1 were log rank p value =0.001, HR= 2.54, (CI 1.39–4.66)), and after adjusting for HER2 expression the HR was 1.95 (CI 1.04–3.68). In METABRIC, the log rank p value, HR, and CI were log rank p value =0.04, HR=1.26, (CI 1.00–1.57)) and after adjusting for HER2 expression, HR = 1.06 (CI 0.84–1.34).

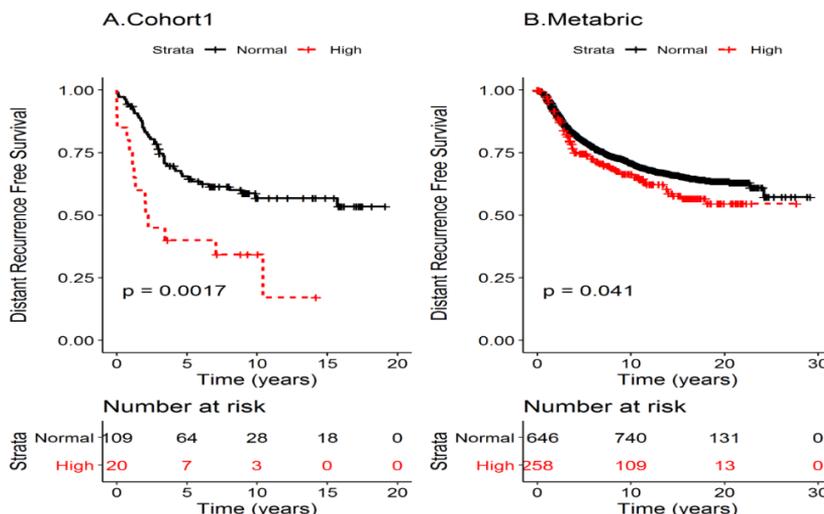


Figure 15. High expression of VMP1 was associated with shorter DRFS time VMP1 mRNA levels within cohort1 and METABRIC were divided into two groups, VMP1 high (\geq mean + 1 SD) and VMP1 normal ($<$ mean + 1 SD). The log rank p values and number of patients at risk are shown in the graphs. DRFS's hazard ratio (HR) in cohort 1 was 2.54 (CI 1.39–4.66), and after adjusting for HER2 expression the HR was 1.95 (CI 1.04–3.68). In METABRIC HR was 1.26 (CI 1.00–1.57) and after adjusting for HER2 expression it was HR = 1.06 (CI 0.84–1.34).

4.2.6 VMP1 mRNA correlates with known drivers of the Chromosome 17q23 locus

RPS6KB1, PPM1D and hsa-miR-21-5p were shown to be drivers at the chromosome 17q23 locus (Holz, 2012; Y. Liu et al., 2018; Noh et al., 2008; Pérez-Tenorio et al., 2011; van der Hage et al., 2004). Thus, we analyzed whether VMP1 expression correlated with these drivers. In METABRIC/EGA (n=1220), VMP1 mRNA correlated with RPS6KB1, PPM1D and hsa-miR-21-5p values as well as has-miR-21-3p. VMP1 mRNA correlated positively only with RPS6KB1 ($r = 0.61$, $p < 2.2e-16$) PPM1D ($r = 0.43$, $p < 2.2e-16$), hsa-miR-21-5p ($r = 0.41$, $p < 2.2e-16$) and as well as hsa-miR-21-3p ($r=0.57$, $p < 2.2e-16$). (Figure 16).

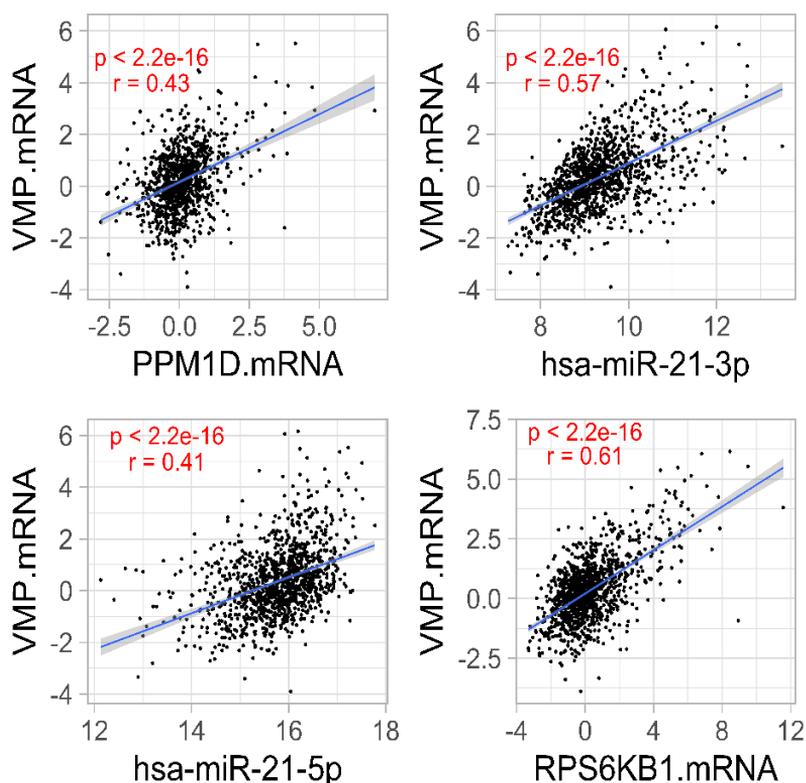


Figure 16. VMP1 mRNA correlates with PPM1D, hsa-miR-21-3p and RPS6KB1
VMP1 mRNA values in METABRIC/EGA (measured via Agilent microarray) correlated with PPM1D mRNA, hsa-miR-21-3p, hsa-miR-21-5p and RPS6KB1 measured by the same technique. A. PPM1D and VMP1 Pearson r value was 0.43, $p < 2.2e-16$. B. hsa-miR-21-3p and VMP1 Pearson r value was 0.57, $p < 2.2e-16$. C. hsa-miR-21-5p and VMP1 Pearson r value was 0.41, $p < 2.2e-16$. D. RPS6KB1 and Pearson r value was 0.61, $p < 2.2e-16$

4.2.7 The effect of elevated expression of VMP1 mRNA on survival is independent of RPS6KB1, PPM1D, miR21 gene expression

As already pointed out, RPS6KB1 is a BC gene (Holz, 2012; Noh et al., 2008; Pérez-Tenorio et al., 2011; van der Hage et al., 2004); and a recent study showed miR21 and PPM1D functionally cooperate with HER2 in breast tumorigenesis (Y. Liu et al., 2018). Due to these reasons and the positive correlation of VMP1 mRNA with expression of RPS6KB1, PPM1D and hsa-miR-21-5p genes, they could have a confounding effect on the effect of VMP1 on survival (Figure 16). Thus, Cox regression analysis was used to check whether RPS6KB1, PPM1D, and miR21 genes cause a confounding effect.

In the TCGA cohort, VMP1, RPS6KB1, PPM1D and hsa-miR-21-5p levels were classified to high and normal groups, based on the mean and SD, as described previously. Overall survival data based on VMP1 high and normal groups were adjusted to high RPS6KB1, PPM1D and hsa-miR-21-5p groups with Cox regression analysis (Table 15). Based on Cox regression analysis, the elevated levels of PPM1D, RPS6KB1 hsa-miR-21-5p genes did not confound the effect of elevated levels of VMP1 on survival data.

Table 15. Adjusted VMP1 cox model to RPS6KB1, PPM1D and hsa-miR-21-5p high groups

17q23 amplicon genes	HR	95% CI	p-value
VMP1 ^{high} (n=53)	2.10	1.09 – 4.04	0.02
VMP1 ^{high} + RPS6KB1 ^{high} (n=53)	3.24	1.54 – 6.82	0.001
VMP1 ^{high} + PPM1D ^{high} (n=51)	2.21	1.02 – 4.82	0.04
VMP1 ^{high} + hsa-miR-21-5p ^{high} (n=82)	2.97	1.15 – 7.72	0.02

4.2.8 HER2 positive patients with high VMP1 had shorter survival

Effect of high VMP1 on DRFS in cohort 1 adjusted for HER2 (reduced to 1.95) remained significant (CI 1.04–3.68), whereas in METABRIC HR was no longer significant (HR 1.06, CI 0.84–1.34) after adjusting for HER2. This suggests confounder effect of HER2 on survival in METABRIC which might be due to different therapy regimens. Association of VMP1 with BCSS and DRFS was assessed in HER2-positive tumors of METABRIC (n=220). METABRIC has enough numbers of HER2 positive tumors and none of patients received trastuzumab and other treatment. Median breast cancer survival of HER2 positive patients in METABRIC with respect to VMP1 high and normal groups was 7.6 years and 12.2 years respectively. VMP1 did not have any effect on BCSS of HER2 positive tumors in METABRIC but it had suggestive weak effect on DRFS (log rank p = 0.085) (Figure 17).

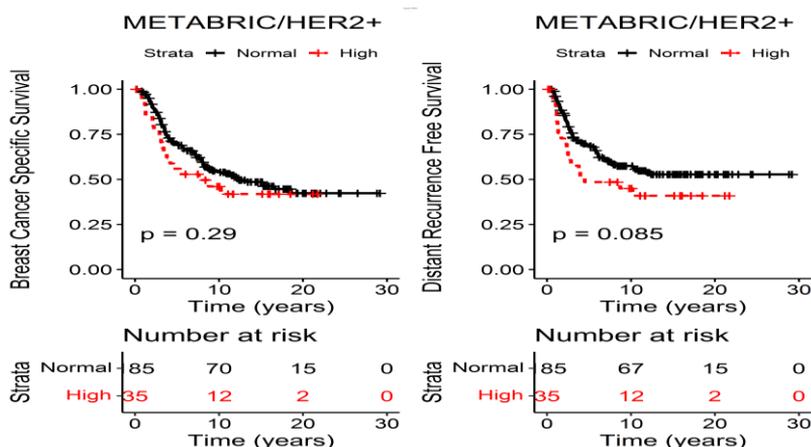


Figure 17. High expression of VMP1 was associated with shorter BCSS and DRFS. In HER2 positive breast cancer patients from METABRIC, patients categorized to high VMP1 (high \geq mean + 1 SD) and normal VMP1 (normal $<$ mean + 1 SD) and analyzed with respect to breast cancer specific survival (BCSS) and distant recurrence free survival (DRFS). The log rank p-values are shown in the figure and the numbers of patients at risk are shown in the table below the graphs.

High expression of VMP1 within HER2 negative patients from METABRIC did not associate with breast cancer specific survival (BCSS) (HR 1.04, CI 0.81–1.35) and distant recurrence free survival (DRFS) (HR 1.04, CI 0.31–0.75) (Appendix figure 5).

VMP1 mRNA was significantly higher in tumors samples than adjacent normal breast tissues. It was higher in VMP1 amplified tumors within cohorts and correlated with VMP1 copy number variation data. VMP1 mRNA was significantly higher in HER2 positive tumors in all four cohorts and high expression of VMP1 associated with shorter BCSS and DRFS in cohort 1 and METABRIC. These results implicate VMP1 in playing a role in development of breast cancer.

4.3 Hsa-miR-21-3p is a marker of poor survival in breast cancer patients (manuscript)

MicroRNA 21 is a known oncomir in various cancers, including breast cancer. It resides at 17q23.1 chromosomal region downstream of 3' UTR of VMP1 gene (Figure 5). The 17q23.1 chromosomal region is a fragile site with many chromosomal rearrangements (such as amplification) that can result in generation of fusion genes. MicroRNA 21 has its own promoter but is processed via a polyadenylation signal different from that of VMP1 (Figure 5).

Our previous study checked whether expression of miR21 confounds the effect of VMP1 on survival. We quantified hsa-miR-21-3p and hsa-miR-21-5p within cohort 1 and noticed differences between 3p and 5p association with clinically pathologic features of tumors. Unlike Hsa-miR-21-5p, for which a role in breast cancer is well described, little is understood about whether there is a role for hsa-miR-21-3p in breast cancer. To examine this, hsa-miR-21-3p was quantified in two breast cancer cohorts and results compared to publicly available database cohorts in TCGA and METABRIC.

4.3.1 Hsa-miR-21-3p was higher in breast tumors

To compare hsa-miR-21-3p expression in normal breast tissues vs tumors, we examined 35 normal breast tissues from cohort 2 that were adjacent to their matched tumors. Tumors expressed significantly more Hsa-miR-21-3p (P value= $4.5e^{-13}$). Likewise, in breast tumors (n=172) and their matched normal breast tissues (n=172) from the TCGA cohort, hsa-miR-21-3p levels were higher in tumors than normal breast tissues (P value < $2e^{-16}$ Figure 18).

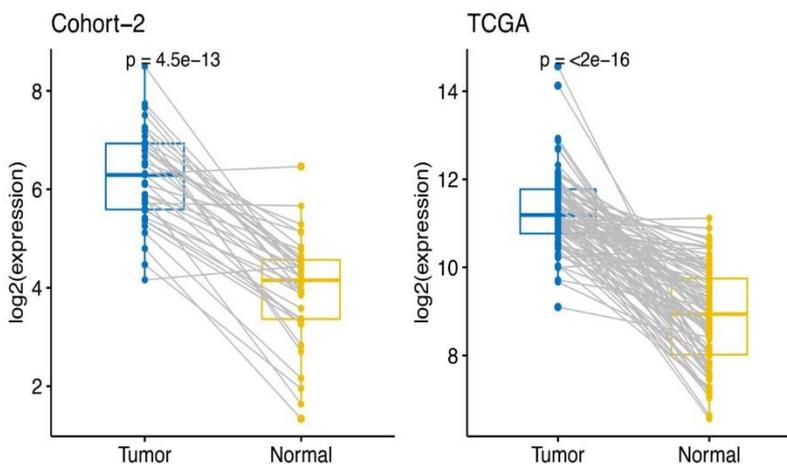


Figure 18. hsa-miR-21-3p levels were higher in breast tumors than paired normal tissues.

Hsa-miR-21-3p levels in breast tumors (n=35) compared to matched normal breast tissues (n=35). Hsa-miR-21-3p data were normalized with log 2 transformation. Paired t-test in R was used to compare hsa-miR-21-3p values in tumors and normal tissue (P value= $4.51e^{-13}$). Hsa-miR-21-3p levels (P value < $2e^{-16}$) compared within tumors (n=172) and matched normal breast tissues (n=172) from the TCGA.

4.3.2 Hsa-miR-21-3p level is highest in MIR21-amplified tumors

To check for an impact of MIR21 CNV on activation of expression of hsa-

miR-21-3p, levels of hsa-miR-21-3p were correlated with MIR21 CNV in cohort 1 and cohort 2. The probes used to measure VMP1 DNA overlap at exon 12 of the VMP1 gene and pri-miR 21, and so potentially detect MIR21 DNA, so VMP1 DNA data was used to analyze the association between hsa-miR-21-3p levels and MIR21 copy number variations. In all four cohorts, copy number variation (CNV) was defined as amplification, gain, neutral and loss based on method used in TCGA data set ("Comprehensive molecular portraits of human breast tumours," 2012). According to this method, amplification refers to 4-6 copies of a DNA region, gain 3-4 copies of the DNA, neutral 2-3 copies and loss ≤ 1 copy of the DNA.

Hsa- miR-21-3p associated with MIR21 CNV within cohort 2 ($p=5.31e^{-05}$), TCGA ($p= 1.3e^{-12}$) and METABRIC ($p < 2e^{-16}$) but not cohort 1 ($P = 0.62$) and was high in amplified tumors (Figure 19), which supports the theory that it is an oncogene.

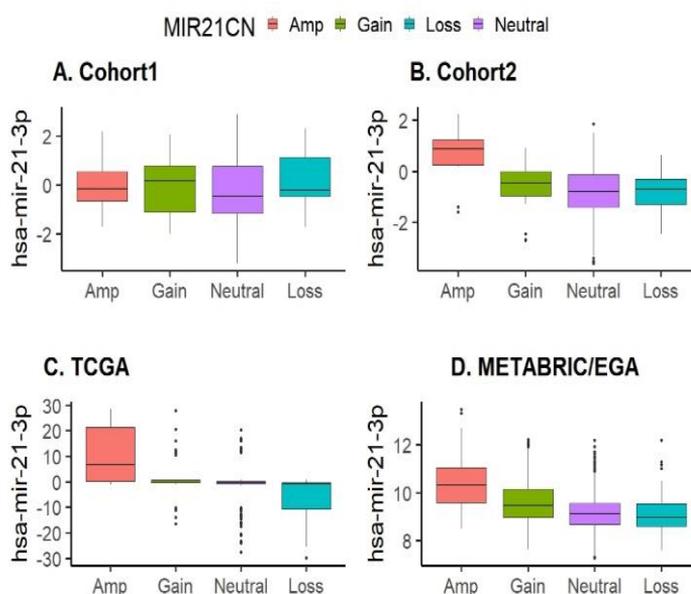


Figure 19. hsa-miR-21-is associated with MIR21 copy number variations. MIR21's DNA levels from A) Cohort 1, B) Cohort 2, C) TCGA and D) METABRIC were categorized as Amplification, Gain, Neutral and Loss according to method used in TCGA (explained in detail in material Methods section). MIR21 CNV frequencies in cohort1 (n=143) were 4.89% amplification (n=7), 9.09% gain (n=13), 72.72% neutral (n=104) and 13.28% loss (n= 19). In cohort 2 (n=273) MIR21 CNV frequencies were 4.21% amplification (n=12), 9.47% gain (n=27), 77.89% neutral (n=222) and 8.42% loss (n= 24). In TCGA (n=816) MIR21 CNV frequencies were 10.27% amplification (n=26), 28.45% gain (n=72), 53.35% neutral (n= 135) and 7.09% loss (n=20). In

METABRIC (n=1980) MIR21 CNV frequencies were 11.55% amplification (n= 141), 12.95% gain (n= 158), 63.93% neutral (n=780) and 11.55% loss (n=141). hsa-miR-21 was compared with ANOVA in R with copy number variations. The p values were cohort 1 P value= 0.62, cohort 2 P value=5.31e⁻⁰⁵, TCGA P value =1.3e⁻¹² and METABRIC P value < 2e⁻¹⁶.

4.3.3 Hsa- miR-21-3p association with clinic pathologic features

To understand whether hsa-miR-21-3p levels could indicate severity of disease the microRNA values were correlated with the tumor's clinical and pathological characteristics in cohorts 1 (appendix table 12), 2 (appendix table 13), TCGA (appendix table 14) and METABRIC/EGA (Table 16).

In the METABRIC cohort hsa-miR-21-3p levels were significantly associated with estrogen receptor (P value=0.004) and HER2 receptor (P value=1.86e-09) and it were higher in ER-negative and HER2-positive tumors. Hsa-miR-21-3p levels were significantly higher in stage 4 (P Value = 0.0005), large tumors (P Value = 0.012), grade 3 tumors (P Value = 6.72e-14) and high cellularity tumors (P Value = 0.02). Hsa-miR-21-3p also was associated with nodal status of tumors (P Value = 0.002) and was higher in node-positive tumors than node-negative tumors (P Value = 0.001). HER2 enriched and Luminal B subtypes had higher levels of hsa-miR-21-3p than others (P Value < 2e-16).

In cohort 1, hsa-miR-21-3p was significantly high in tumors with metastasis (p=0.02). In cohort 2, hsa-miR-3p was high in ER-negative tumors (p=0.03), HER2-positive tumors (p=0.003) and grade 3 tumors (p=0.02). In TCGA it was only high in HER2-positive tumors (p=0.001).

Table 16. hsa-miR-21-3p association with clinically pathologic features in METABRIC/EGA.

n	1220	miR21-3p mRNA	p-value
		median (25 and 75%)	
Age			0.941
< 50	279	-0.011 (-0.467, 0.653)	
≥ 50	941	0.005 (-0.427, 0.557)	
Estrogen receptor			0.004
Negative	281	0.190 (-0.320, 0.718)	
Positive	939	-0.031 (-0.465, 0.526)	
Progesterone receptor			0.143
Negative	581	0.064 (-0.381, 0.642)	
Positive	639	-0.053 (-0.480, 0.523)	
HER2 status			1.86e ⁻⁰⁹
Negative	1067	-0.039 (-0.473, 0.500)	
Positive	153	0.448 (-0.110, 1.353)	
Tumor stage			0.0005
1	364	-0.097 (-0.506, 0.374)	
2	588	0.079 (-0.372, 0.653)	
3	98	0.020 (-0.311, 0.712)	
4	10	0.656 (0.230, 1.010)	
Unknown	158		
Tumor size (mm)			0.012
≤ 20	527	-0.023 (-0.448, 0.486)	
> 20	680	0.047 (-0.410, 0.662)	
			6.72e ⁻¹⁴
Histologic Grade			
1	106	-0.255 (-0.521, 0.132)	
2	494	-0.117 (-0.581, 0.330)	
3	620	0.211 (-0.324, 0.817)	
Cellularity			0.023
Low	137	-0.011 (-0.598, 0.529)	
Moderate	447	-0.036 (-0.453, 0.519)	
High	586	0.058 (-0.399, 0.676)	
Unknown	50		
Nods			0.001
Negative	623	-0.085 (-0.521, 0.517)	
Positive	551	0.058 (-0.351, 0.685)	
Unknown	46		

Nodal status			0.002
N0	623	-0.085 (-0.521, 0.517)	
N1	358	0.054 (-0.359, 0.561)	
N2	127	0.231 (-0.370, 1.005)	
N3	66	0.052 (-0.267, 0.804)	
Unknown	46		
Histology subtype			4.78e ⁻⁰⁷
Ductal/NST	922	0.073 (-0.366, 0.666)	
Lobular	96	-0.347 (-0.785, 0.051)	
Medullary	15	-0.027(-0.306, 0.726)	
Metaplastic	0		
Mixed	139	-0.023 (-0.613, 0.313)	
Mucinous	12	-0.649 (-0.973, -	
Other	7	0.629 (-0.826, 1.734)	0.454)
Tubular/cribriform	15	-0.291 (-0.695, 0.599)	
Unknown	14		
Subtype PAM50			< 2e ⁻¹⁶
Basal	118	-0.076 (-0.479, 0.469)	
Claudin-low	148	0.249 (-0.20, 0.719)	
Her2	109	0.325 (-0.186, 1.032)	
LumA	385	-0.144 (-0.565, 0.238)	
LumB	294	0.242 (-0.344, 1.071)	
NC	4	-0.157 (-0.,573,	
Normal	102	-0.312 (-0.766, 0.454)	0.356)
Unknown	60		
3-Gene Classifier Subtype			< 2e ⁻¹⁶
ER-/HER2-	196	0.089 (-0.338, 0.545)	
ER+/HER2- High Prolif	394	0.106 (-0.402, 0.780)	
ER+/HER2- Low Prolif	411	-0.163 (-0.594, 0.261)	
HER2+	127	0.464 (-0.059, 1.380)	
Unknown	92		

*Two stage 0 tumors were removed from the analysis. The table shows the median and the 25th and 75th percentiles. The p-value is calculated using a t-test or ANOVA. Hsa-miR-21-3p values transformed with z-scores.

4.3.4 Hsa-miR-21-3p levels were higher in HER2-positive tumors than HER2-negative tumors

Hsa-miR-21-3p was associated with HER2 status of tumors determined with

immunohistochemistry in cohort 2 ($P = 0.003$), TCGA (P value = 0.001) and METABRIC (P value = $1.86e^{-09}$) cohorts and was higher in HER2-positive tumors than HER2-negative tumors (Figure 20).

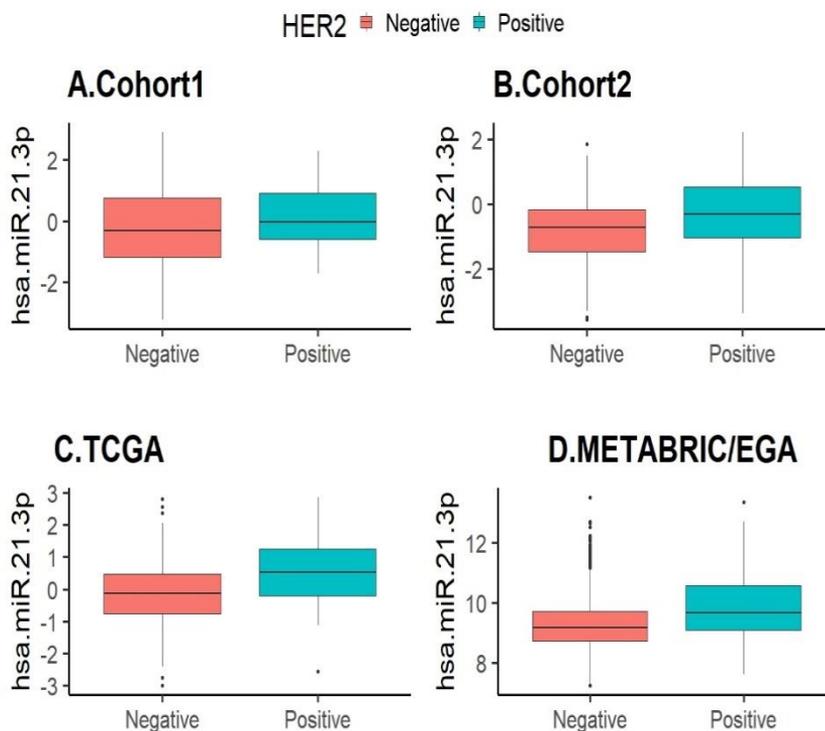


Figure 20. Hsa-miR-21-3p was high in HER2 positive tumors. Hsa-miR-21-3p levels within A) Cohort 1, B) Cohort 2, C) TCGA and D) METABRIC were compared with HER2 status of tumors determined by immunohistochemistry. Student t-test in R was used to compare 2 groups. Based on t-test results, p values within cohorts were cohort 1 (P Value = 0.28), cohort 2 (P Value = 0.003), TCGA (P Value = 0.001) and METABRIC (P Value = $1.86e^{-09}$). Hsa-miR-21-3p data were normalized with log 2 transformation within cohorts 1 and 2.

4.3.5 Hsa-miR-21-3p associated with histological grade of breast tumors

Hsa-miR-21-3p levels correlated with histological grade of tumors within cohorts 1, 2, and METABRIC (Figure 21). For the TCGA data histological grade was not available. METABRIC was the only cohort in which hsa-miR-21-3p was associated with the histological grade of tumors (P value = $6.74e^{-}$

¹⁴). In the METABRIC cohort, there were 106 grade 1, 494 grade 2 and 620 grade 3 tumors. Hsa-miR- 21-3p levels were highest in grade 3. Based on the Nottingham Grading System, grade 3 breast tumors are poorly differentiated with frequent mitosis and no tubule formation.

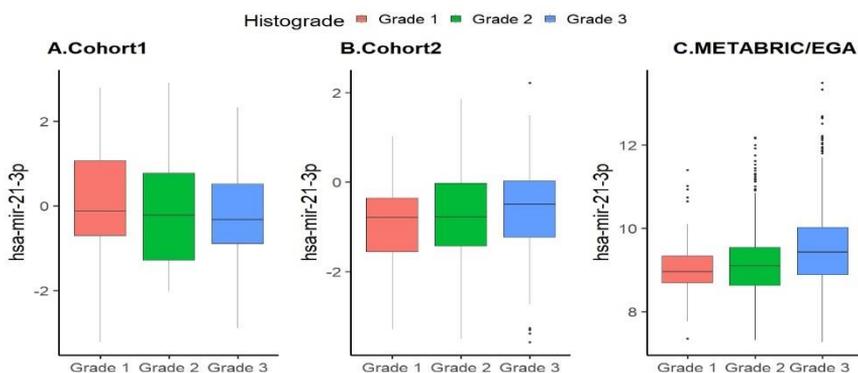


Figure 21. Hsa-miR-21-3p was high in grade 3 tumors within cohort 2 and METABRIC

Hsa-miR-21-3p levels within A) Cohort 1, B) Cohort 2, and C) METABRIC were compared with histograde of tumors. ANOVA in R was used for statistical analysis. Hsa-miR-21-3p data were normalized with log 2 transformation within cohorts 1 and 2. For cohort 1, P value= 0.77); for cohort 2 P value=0.08; and for METABRIC P value=6.74e¹⁴. Histograde data for TCGA is not available.

4.3.6 High expression of hsa-miR-21-3p as a marker of short survival

Disease free survival (DFS) refers to a period from initiation of treatment until progression of disease from any cause and is a valuable measure of benefit of cancer patients from adjuvant treatment after surgery and radiotherapy (U.S. Food and Drug Administration December 2018). Due to this, levels of hsa-miR-21-3p within cohort 1, cohort 2 and TCGA based on median divided to two high and low groups and associated with DFS data. In cohort 1, HR and CI were 1.89 and (1.18-3.03), respectively, and log rank p value was 0.007. In cohort 2 HR was 1.52, CI was (0.97-2.36) and log rank p value was = 0.06. In TCGA, elevated levels of hsa-miR-21-3p did not associate with DFS (Figure 22). In METABRIC, data for DFS was not available. With respect to this analysis, high miR21-3p levels were associated with shorter DFS only in cohort 1.

HER2 is a strong oncogene and hsa-miR-21-3p expression levels were high in HER2-positive tumors. For this reason, effect of HER2-positive tumors

was checked with Cox regression analysis on effects of high hsa-miR-21-3p levels on DFS in cohort 1, HER2 did not confound high hsa-miR-21-3p effect of DFS. Adjusted HR and CI for HER2 in cohort 1 was (HR: 1.72 CI: 1.07-2.78). Since reports are inconsistent regarding the role of hsa-miR-21-3p in metastasis in tumor types (Báez-Vega et al., 2016; Jiao et al., 2017; Lo, Tsai, & Chen, 2013; Pink et al., 2015), its levels in METABRIC were analyzed with respect to distant metastasis free survival data (DMFS). In METABRIC, patients carrying tumors with elevated levels of hsa-miR-21-3p had shorter DMFS time compared to patients carrying tumors with low levels (log rank p value=0.002; Figure 23 B). In patients carrying tumors with elevated levels of hsa-miR-21-3p, DMFS HR was 1.36 and the CI was 1.11-1.67 (after adjusting for HER2 expression HR = 1.28; CI :1.04-1.58). BCSS data indicates death due to BC. To analyze whether hsa-miR-21-3p had an effect on BCSS, levels of hsa-miR-21-3p in METABRIC were divided into high and low groups based on the median value and associated with BCSS data (Figure 23 A). Breast cancer specific HR and CI within METABRIC were 1.394 and 1.146-1.695, respectively. The log rank p-value was 0.0008.

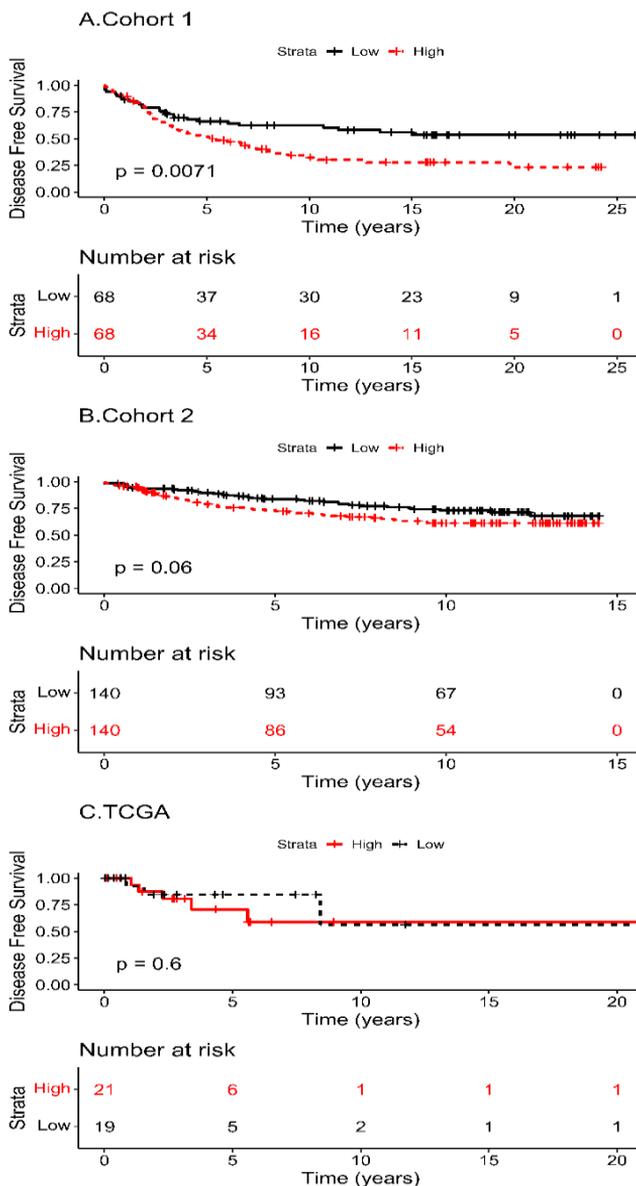


Figure 22. High expression of hsa-miR-21-3p was associated with shorter disease-free survival.

Hsa-miR-21-3p levels within cohort 1, cohort 2 and TCGA divided into 2, high and low groups based on median. The number of patients at risk are shown below the graphs. The disease-free survival's hazard ratio (HR) in cohort 1 was 1.89 (CI: 1.18-3.03); and after adjusting for HER2 the HR was 1.72 (CI :1.07-2.78). In cohort 2 HR was 1.52 (CI: 0.97-2.36). TCGA gives DFS values for 240 of the 256 patients with miR21-3p measurements, subtracting seven from the low category and nine from the high category.

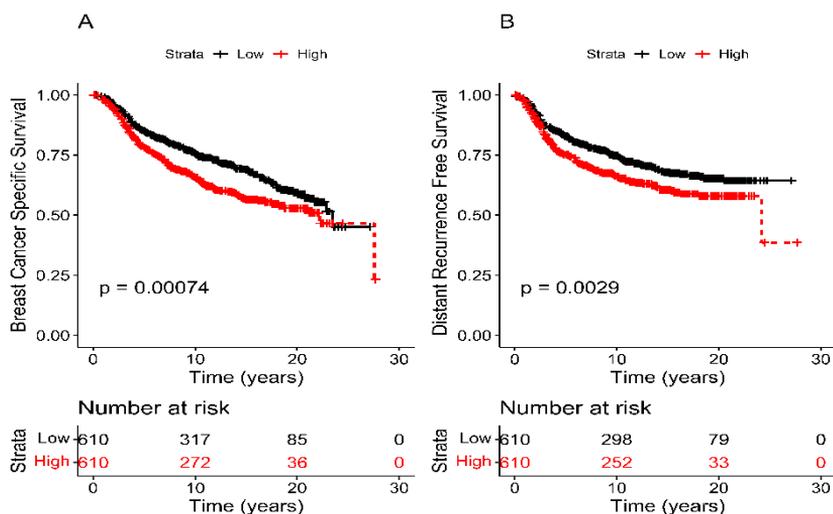


Figure 23. High expression of hsa- miR-21-3p was associated with shorter BCSS and DRFS within METABRIC/EGA.

Hsa- miR-21-3p levels within METABRIC divided into two groups, high and low based on median, and the number of patients at-risk is shown in the graphs, A. hazard ratio (HR) for BCSS was 1.394, (CI :1.146-1.695). B. HR for DRFS was 1.36 (CI: 1.11-1.67).

4.3.7 Hsa-miR-21-3p affected survival independently of other clinically pathological features and neighboring genes

ER negativity, HER2 positivity, high grade and positive lymph node status of breast tumors results in shorter survival of breast cancer patients, as shown in numerous studies over the years. Elevated levels of hsa-miR-21-3p were seen in ER-negative, HER2- positive, lymph-positive and high-grade tumors within the METABRIC/EGA cohort. RPS6KB1 (Holz, 2012; Noh et al., 2008; Pérez-Tenorio et al., 2011; van der Hage et al., 2004), hsa-miR-21-5p , PPM1D (Y. Liu et al., 2018) and VMP1 (Amirfallah et al., 2019) have been shown as potential breast cancer genes and positively correlate with hsa-miR-21-3p.

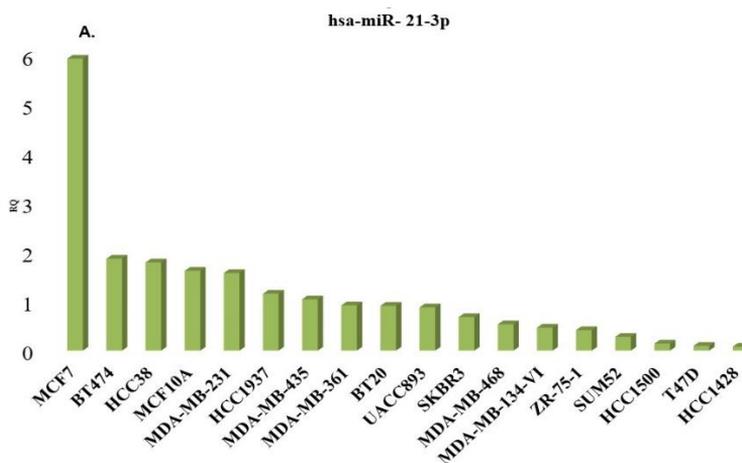
The METABRIC/EGA cohort Cox model with high hsa-miR-21-3p was adjusted for clinical factors like estrogen, HER2, node status, grade and expression of other neighbor genes of hsa-miR-21-3p like VMP1, RPS6KB1, PPM1D and hsa-miR-21-5p. All the coefficients describing HRs, CIs and p-values of Cox regression analysis are shown in Table 17. The biggest confounding effect was from high grade tumors (grade 2 and grade 3), yet the p value was still significant (p= 0.02). Based on these data, hsa-miR-21-3p appears as a marker of short survival in breast cancer patients.

Table 17. Adjusted hsa- miR-21-3p cox model to neighbor genes expression and clinic-pathological features

	HR	CI	p-value
hsa- miR-21-3p ^{High}	1.39	1.15-1.7	0.00078
hsa- miR-21-3p ^{High} + ER ⁻	1.36	1.11-1.65	0.002
hsa- miR-21-3p ^{High} + HER2 ⁺	1.31	1.07 - 1.6	0.007
hsa- miR-21-3p ^{High} + Node ⁺	1.34	1.09 - 1.63	0.004
hsa- miR-21-3p ^{High} + Grade ^{2&3}	1.25	1.02 - 1.52	0.02
hsa- miR-21-3p ^{High} +VMP1 ^{High}	1.44	1.16 - 1.77	0.0006
hsa. miR.21.3p ^{High} +RPS6KB1 ^{High}	1.39	1.13 - 1.71	0.001
hsa- miR-21-3p ^{High} +PPM1D ^{High}	1.37	1.13-1.68	0.001

4.3.8 Hsa-miR-21-3p levels within breast cancer cell lines

A total of 18 breast cell lines was used to quantify hsa-miR-21-3p levels; 17 of them were cancer cell lines and one a normal breast cell line. Seven cell lines were ER⁻/HER2⁻, two ER⁻/HER2⁺, five ER⁺/HER2⁻ and three ER⁺/HER2⁺. MCF7 cells had higher levels of hsa-miR-21-3p than other cell lines used in this experiment. Unlike breast tumors from METABRIC/EGA in which there was a significant difference between levels of hsa-miR-21-3p within ER⁻/HER2⁻ (n=202), ER⁻ /HER2⁺ (n=79), ER⁺/HER2⁻ (n=865) and ER⁺/HER2⁺ (n=74) groups (p=1.05e⁻¹¹), there was not a significant difference between levels of hsa-miR-21-3p within ER⁻/HER2⁻, ER⁻ /HER2⁺, ER⁺/HER2⁻ and ER⁺/HER2⁺ groups among the cell lines (Figure 24), which is due to low number of cell lines in each category. Hsa-miR-21-3p levels in cell lines similar to tumors were high in ER⁺/HER2⁺ and in ER⁻/HER2⁺ groups



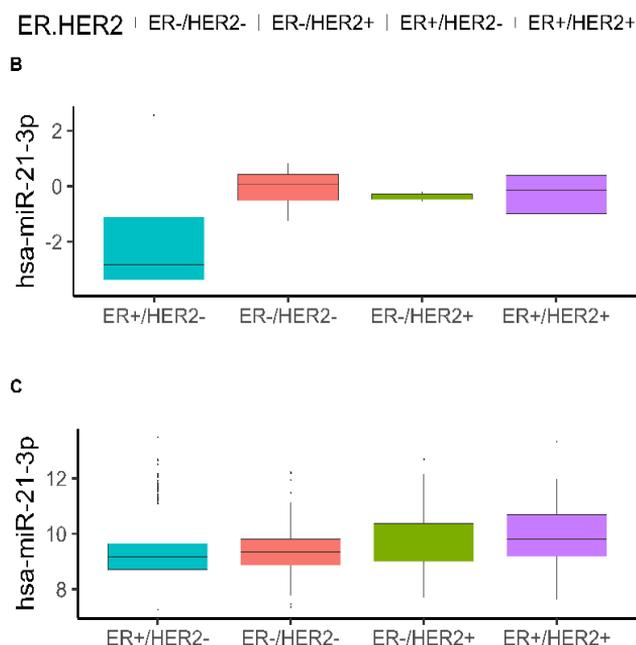


Figure 24. Levels of hsa-miR-21-3p were high only in ER⁺/HER2⁺ tumors. (A). hsa- miR-21-3p levels were quantified in breast cancer cell lines (n=17) and one normal breast cell line. (B) Cell lines based on expression of ER and HER2 were categorized to ER⁻/HER2⁻ (n=7), ER⁻/HER2⁺ (n=2), ER⁺/HER2⁻ (n=5), ER⁺/HER2⁺ (n=3) and compared with respect to hsa-miR-21-3p levels with Anova test, p value (NS). (C). Tumors in METABRIC/EGA based on expression of ER and HER2 were categorized to ER⁻/HER2⁻ (n=202), ER⁻/HER2⁺ (n=79), ER⁺/HER2⁻ (n=865), ER⁺/HER2⁺ (n=74) and compared with respect to hsa-miR-21-3p levels with Anova test, P value ($1.05e^{-11}$).

In summary, hsa-miR-21-3p was higher in breast tumors than matched adjacent normal breast tissues, it was associated with MIR21 gene copy number variation, and was higher in amplified tumors, it was higher in ER negative, stage 4, tumors larger than 20mm, grade 3, node positive and the HER2 enriched subtypes. Tumors with elevated levels of hsa-miR-21-3p had shorter DFS and BCSS, even after adjusting for ER, HER2, node, grade and hsa-miR-21-3p neighbor gene levels. Taking these data regarding association of hsa-miR-21-3p with clinic pathological features and survival data of breast patients into consideration, these results suggest hsa-miR-21-3p has role in the development of breast cancer.

4.4 Study of VMP1 function in cell lines (unpublished data)

Our previous results regarding VMP1's role in breast tumors showed that VMP1 is highly expressed in HER2 positive tumors and VMP1 expression is associated with poor prognosis in breast cancer patients, particularly in patients with HER2 positive tumors. VMP1 has a role in autophagy and formation of tight junctions through its interaction with ZO-1 (Sauermaun et al., 2008) . VMP1's role in breast cancer has not been well studied but the above-mentioned information prompted us to propose that VMP1 could have a role in the tumorigenesis of HER2 positive breast cancer cells and that VMP1 could play a role in drug resistance in HER2 positive cells through its function in autophagy.

4.4.1 Characteristics of cell lines

To explore these two hypotheses, suitable breast cancer cell lines had to be identified. VMP1 and ERBB2 copy number and mRNA expression data were retrieved from Cosmic/ cell lines project (https://cancer.sanger.ac.uk/cell_lines) for 50 breast cancer cell lines. The Cosmic/ cell lines project provides mutation profiles for over 1000 cell lines used in cancer research. To select cell lines to study the role of VMP1, we considered factors such as, expression of HER2 in HER2 positive tumors, the VMP1 CN, VMP1 expression, VMP1 fusion, and doubling time (DT). We also examined response to trastuzumab since autophagy plays a role in resistance of tumors to trastuzumab (Table 18). BT-474 and MDA-MB-361 cell lines were chosen for studying VMP1's function in cell lines because they both are HER2 and ER positive; and VMP1 and ERBB2 are both amplified and overexpressed in them. BT-474 is an epithelial cell line from a solid, invasive ductal carcinoma of the breast and MDA-MB-361 is a metastatic breast cancer cell line from brain. BT474 does not have any VMP1 fusion, but MD-MB-361 has VMP1 fusion with BRIP1 gene (but this fusion does not have an ORF). DT of BT-474 is 3.7 days and MDA-MB-361 is 3 days. BT-474 cells have the highest trastuzumab response 13.5%-34.3% and the trastuzumab response of MDA-MB-361 is not available.

Table 18. VMP1 expression status within HER2 positive BC cell lines

Cell line	ERBB2 mRNA	ERBB2 CNV	ERBB2 Number of copies	ER	PR	HER2	VMP1 mRNA	VMP1 CNV	Number of copies VMP1	VMP1 ORF fusion	Trastuzumab response % 3d, 5d	DT (Day)
HCC1419	over	gain	14	+	-	+	-	-	-	-	NA	7
AU565	over	gain	14	-	-	+	-	-	-	-	NA	1.5
UACC-893	over	gain	14	-	-	+	-	-	-	-	NA	6.3
UACC-812	over	gain	12	+	-	+	-	-	-	-	NA	4.1
HCC1954	over	gain	14	-	-	+	-	-	-	-	(9.9%-9.9%)	1.9
EFM-192A	over	gain	14	+	+	+	over	-	-	-	NA	9
ZR-75-30	over	gain	14	+	-	+	-	-	-	-	NA	4.5
HCC1569	over	gain	13	-	-	+	-	-	-	-	(13%-13%)	NA
MDA-MB-330	over	gain	14	-	-	+	-	-	-	-	NA	NA
HCC202	over	gain	11	-	-	+	over	-	-	-	(6.9%-2%)	8.3
HCC2218	over	gain	14	-	-	+	over	Gain	14	-	NA	5
BT-474	over	gain	14	+	+	+	over	Gain	10	-	(13.5%-34.3%)	3.7
OCUB-M	over	gain	12	NA	NA	+	-	-	-	-	NA	Na
MDA-MB-361	over	gain	9	+	-	+	over	Gain	8	-	NA	3
JIMT-1	over	gain	12	NA	NA	+	-	-	-	-	(6.6%-3.1%)	1.6
SKBR3	over	gain	NA	-	-	+	-	-	-	-	(26.7%-N/A)	2.3
KPL4	over	gain	NA	-	-	+	-	-	-	-	(14.5%-N/A)	1.25

4.4.2 VMP1 expression was high in ER⁺/HER2⁺ cell lines

For follow-up data of tumor samples and to confirm VMP1 overexpression in BT-474 and MDA-MB-361 cell lines, levels of VMP1 were quantified in eight breast cancer cell lines from our laboratory archive (Figure 25A). Data from 46 breast cancer cell lines from the Broad Institute Breast cancer cell lines (Figure 25B) was compared to VMP1 expression in breast tumors from the METABRIC cohort (Figure 25C). The VMP1 level in cell lines from our lab (n=8) and Broad Institute Breast cancer cell lines (n=46) and tumors from METABRIC (n=1980) was higher in ER positive and HER2 positive cell lines from Broad Institute ($p=0.01$) and tumors ($p<2e-16$; Figure 25). This data in cell lines was supportive of tumor data and made BT474 and MDA-MB-361 cell lines strong candidate lines for our next cell-based experiments.

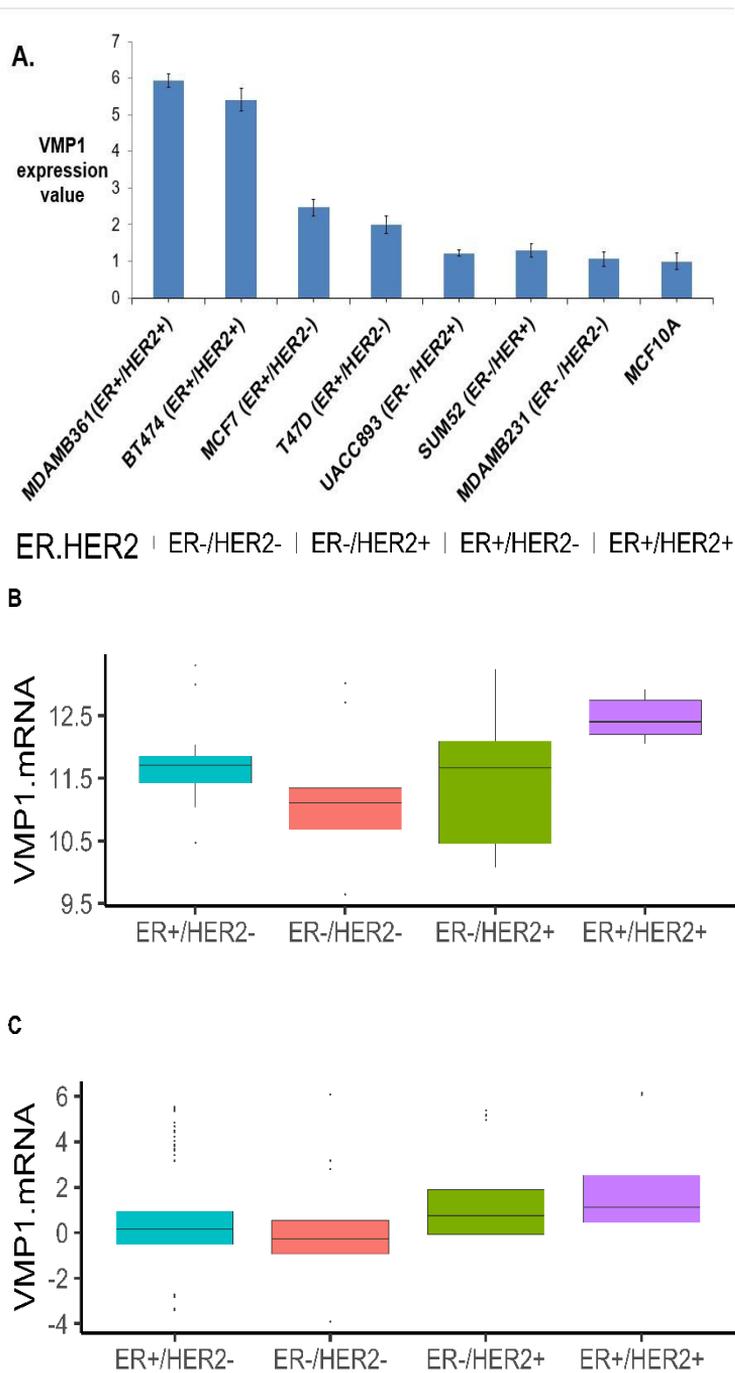
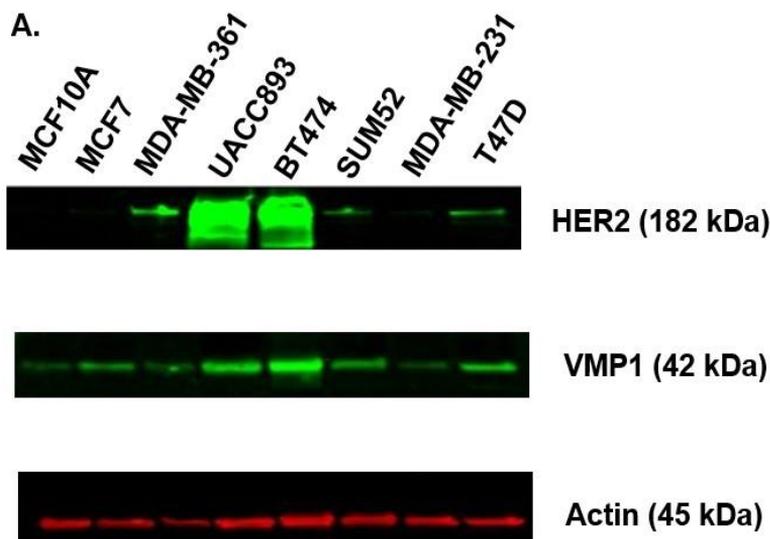


Figure 25. In cultured cells, VMP1 mRNA levels were highest in ER+/HER2+ BC lines, consistent with data in BC tumors.

A: VMP1 mRNA levels were quantified in 8 breast cancer cell lines from our laboratory by using Taqman Gene Expression Assays spanning exons 10-11 (E10-11). B: VMP1 mRNA data from 46 breast cancer cell lines from the Broad Institute Breast cancer. VMP1 mRNA was quantified with Affymetrix microarray. Based on estrogen receptor and HER2 receptor expression, the cell lines were categorized into four groups: ER⁻/HER2⁻ (20), ER⁻/HER2⁺ (9), ER⁺/HER2⁻ (12), and ER⁺/HER2⁺ (5). C: VMP1 mRNA data from METABRIC was used. In this cohort, VMP1 mRNA was quantified with Illumina Human v3 microarray technology. Based on estrogen receptor and HER2 receptor expression, tumors were categorized to four groups: ER⁻/HER2⁻=335, ER⁻/HER2⁺=139, ER⁺/HER2⁻=1398, ER⁺/HER2⁺=108. Both in cell lines and tumors, ANOVA in R was used for statistical analysis (p value for cell lines = 0.01 and for tumors <2e-16).

4.4.3 VMP1 protein was higher in ER⁺/HER2⁺ cell lines

In the eight cell lines our lab had isolated, VMP1 and HER2 protein expression was compared by immunoblotting, showing the BT-474 and UACC893 cells expressed the most protein (Figure 26). Analysis of mRNA levels also detected high expression of VMP1 mRNA in BT-474 cells; in contrast, UACC-893 cells did not express among the highest level of VMP1 mRNA. In MDA-MB-361 cells, mRNA expression of VMP1 was high, but at the protein level was low, possibly a manifestation of the VMP1:BRIP1 fusion in MDA-MB-361 cells.



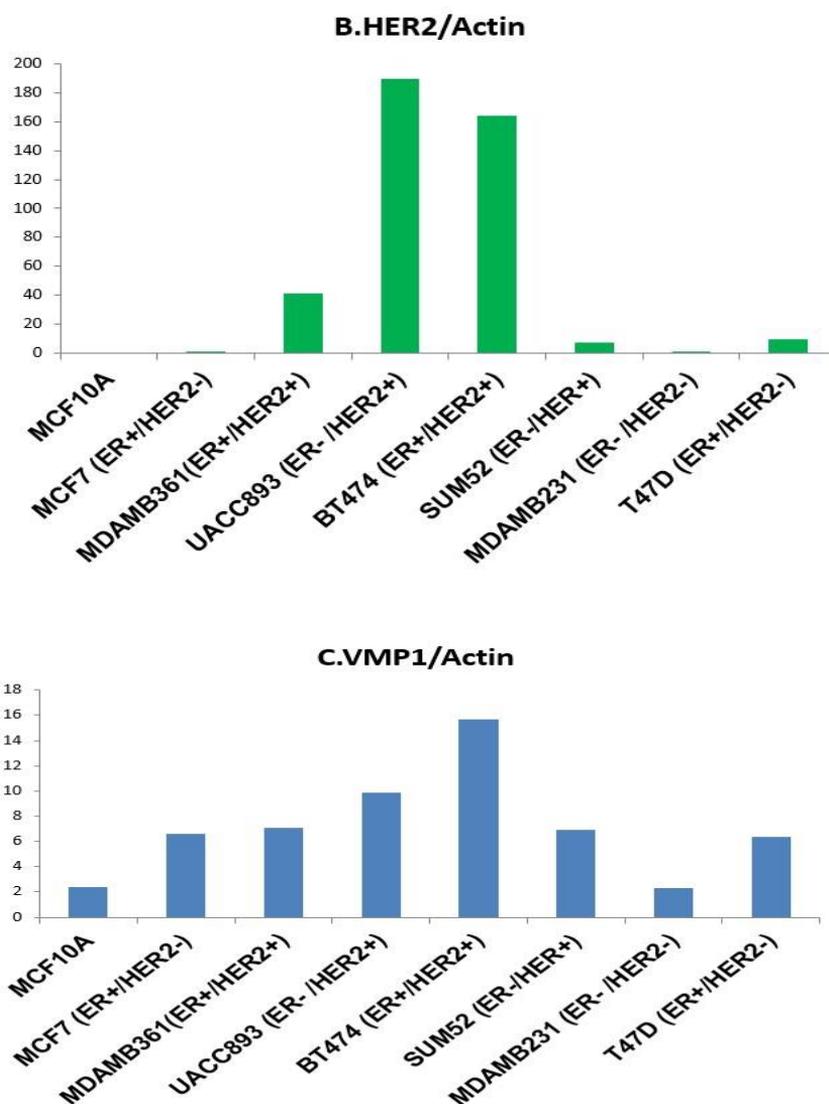


Figure 26. BT-474 cell lines express the most VMP1 protein.

A: Immunoblotting of 8 breast cancer cell lines with antibodies to VMP1 and HER2. Anti-actin was used as a loading control. B: HER2 protein quantification among 8 cell lines. C: VMP1 protein quantification among 8 cell lines. Both HER2 and VMP1 quantifications were normalized to actin expression among cell lines.

4.4.4 Silencing of VMP1 and ERBB2 were optimized in BT474 and MDA-MB-361 cells

Since both the BT474 and MDA-MB-361 cell lines expressed elevated levels of HER2 and VMP1, we used cell-based assays to target VMP1 expression via siRNA. To optimize our analysis conditions, the half-life of VMP1 and HER2 proteins was tracked in BT-474 and MDA-MB-361 cell lines; at 24-hour time points cells were lysed for mRNA and protein extraction. Because the doubling time of the two cell lines is long (3 to 4 days) this experiment ran for 144 hours (six days). Extracted mRNAs and proteins were probed via q-PCR and immunoblotting to track expression of VMP1 and ERBB2/HER2. In BT-474, the maximum knock down of VMP1, both at the mRNA and protein level, was achieved 96 hours after adding siRNAs; maximum knockdown of ERBB2 was achieved 72 hours after adding siRNA at both the mRNA and protein levels. In MDA-MB-361 cells, knockdown of VMP1 and ERBB2 peaked 72 hours after adding siRNAs, at both the mRNA and protein levels. All treatment conditions with VMP1 and ERBB2 siRNAs were compared to a control siRNA with a scrambled sequence (Figure 27).

To maximize knockdown efficiency, two different siRNA assays and their combination for each of VMP1 (s37755, s37756) and ERBB2 (s611, s613) genes were used. Seventy-two hours after adding siRNA to BT-474 and MDA-MB-361 cells, lysates from each treatment condition were analyzed by immunoblot, probing with VMP1 and HER2 antibodies (Figure 28). In BT474 cells, 92% knockdown was achieved using the s37756 assay for the VMP1 gene and s611 for the ERBB2 gene. In MDA-MB-361 cells, 67% and 82% knockdown were achieved by using siVMP1 (s37756) and siERBB2 (s613), respectively. All treatment conditions in both cell lines were confirmed with Taqman gene expression assays for both VMP1 and ERBB2 gene with q-PCR (Appendix figure 3). The s37755 assay spans the junction of exon 8 and 9 of VMP1. The s37756 assay spans the junction of exons 5 and 6 of VMP1 transcripts (www.thermofisher.com). S611 and s613 are both validated assays for ERBB2 silencing. S611 and s613 assays span the junction of exon 3 and exon4 and exons 22 and 23 of ERBB2 transcripts, respectively (www.thermofisher.com; refer to data shown in Figure 27 and Appendix Figure 3). The best siRNA assay for silencing VMP1 in BT474 and MDA-MB-361 cell lines was s37756, and s611 in BT474 cell line; and for knockdown of ERBB2 in MDA-MB-361 cells, s613 was optimal. A combination of s37755 and s37756 silenced VMP1; and s611 and s613 was not better at silencing ERBB2. All experiments were performed with using s37756 for silencing VMP1 and s611 and s613 for silencing ERBB2 in BT474 and MDA-MB-361 cells, respectively.

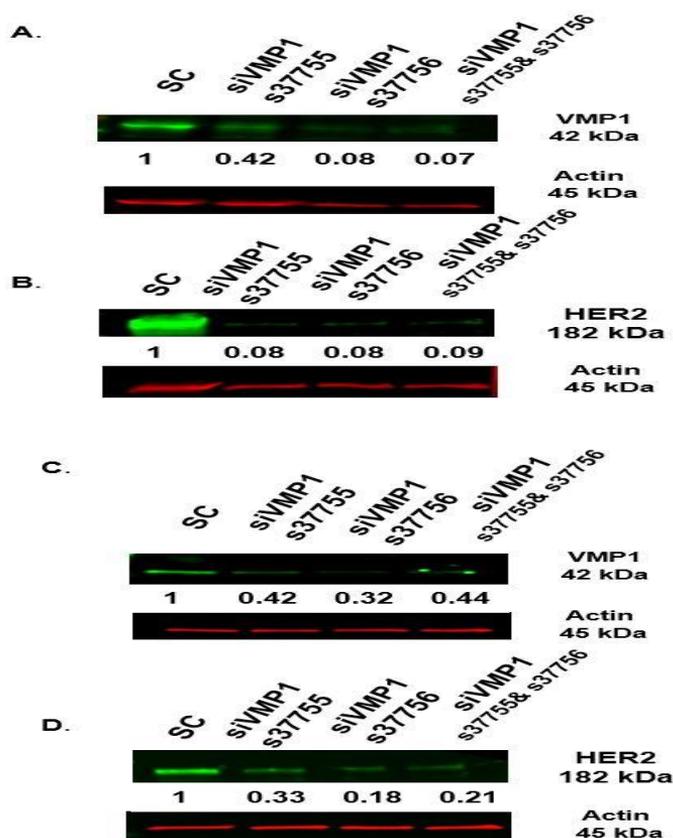


Figure 27. Validation of VMP1 and ERBB2 knockdown.

A, B, C, D: Immunoblotting confirms VMP1 and ERBB2 gene silencing in A9 and B) BT474 and C) and D) MDA-MB-361 cell lines by using two different siRNA assays for VMP1 (s37755) and (s37756) and their combinations (s37756+s37756); and (s611) and (s613) and their combinations (s611+s613) for ERBB2 gene. Seventy-two hours after adding siRNAs, protein and mRNA was extracted from cells. Both HER2 and VMP1 quantifications shown under each figure. Both genes' quantifications were normalized based on actin expression. A and B represent data from the BT474 cell line and C and D from the MDA-MB-361 line.

4.4.5 Effect of VMP1 knockdown on cell proliferation and survival

Uncontrolled growth and evasion from apoptosis are two hallmarks of cancer (Hanahan & Weinberg, 2011). Loss of apoptosis allows survival of cancer cells and increases their invasion capacity during tumor development (Pfeffer & Singh, 2018). Cell proliferation assays can directly measure cell division events, so we used proliferation and apoptosis assays to determine any effect of knocking down of VMP1 on cell proliferation and survival. The

assays were set up such that VMP1 and ERBB2 were silenced separately, or simultaneously, to analyze whether reducing VMP1 expression influenced its own or in the context of ERBB2.

4.4.5.1 VMP1 levels do not correlate with proliferation of BT474 cells

The Incucyte® Live-Cell Analysis System was used for real time monitoring and counting of BT474 and MDA-MB-361 cells following treatment with siRNAs targeting VMP1 and ERBB2 genes. The advantage of the Incucyte® Live-Cell Analysis System is real time monitoring of cells inside incubator and taking pictures from cells every 2 hours and performing data analysis at the end of experiment. The MDA-MB-361 cell line was eliminated from proliferation experiments due to inconsistency results in each repeat. Since the doubling time of BT474 cells was 3.5 days, cells were monitored for seven days to allow for two cell cycles. siERBB2 treatment conditions were used as a positive control. This experiment was done in triplicate and, each time, VMP1 and ERBB2 silencing efficiency were checked with Western blot 96 and 72 hours after adding siVMP1 and siERBB2, respectively. Here, VMP1 knockdown did not affect proliferation of BT474 cells (Figure 28), whereas silencing of ERBB2, or ERBB2 and VMP1 together had similar effects and slowed proliferation of cells. Still, for these cases, it was clear the effect on proliferation was from silencing ERBB2.

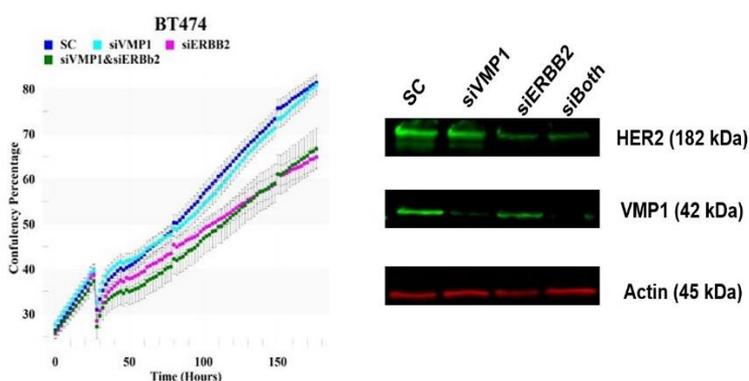


Figure 28. VMP1 silencing in BT474 cells did not affect proliferation.

A. BT474 cells treated with siRNAs targeting VMP1 and ERBB2 genes vs SCRAMBLE, each performed as ten replicates. Cells in each treatment condition were counted every two hours with IncuCyte® Live Cell Analysis and grown for seven days. Cells treated with SC were used as negative control and with siERBB2 as positive control. For checking cooperativity of VMP1 and ERBB2 genes, siERBB2 and siVMP1 were added to the cells simultaneously. B. Immunoblot probed with antibodies to VMP1 and ERBB2 to confirm the knockdown of VMP1, ERBB2 and both.

4.4.5.2 Effects of silencing of VMP1 on apoptosis in BT474 cells

In cancer, apoptosis is a key mechanism of inhibition, so the effect of VMP1 silencing on apoptosis was tracked in BT-474 cells. Ninety-six hours after silencing VMP1, early apoptosis, apoptosis, and necrosis was induced in 7.24%, 1.74% and 6.2% of cells, respectively (as compared to controls; see Figure 29). We controlled for the fact that the transfection reagents alone cause apoptosis with the scrambled control. That ERBB2 gene silencing induces apoptosis is well known (Carpenter & Lo, 2013; Faltus et al., 2004) so SiERBB2 was used as positive control. Here, targeting VMP1 did not affect cellular apoptosis either in the presence or absence of simultaneously targeting ERBB2.

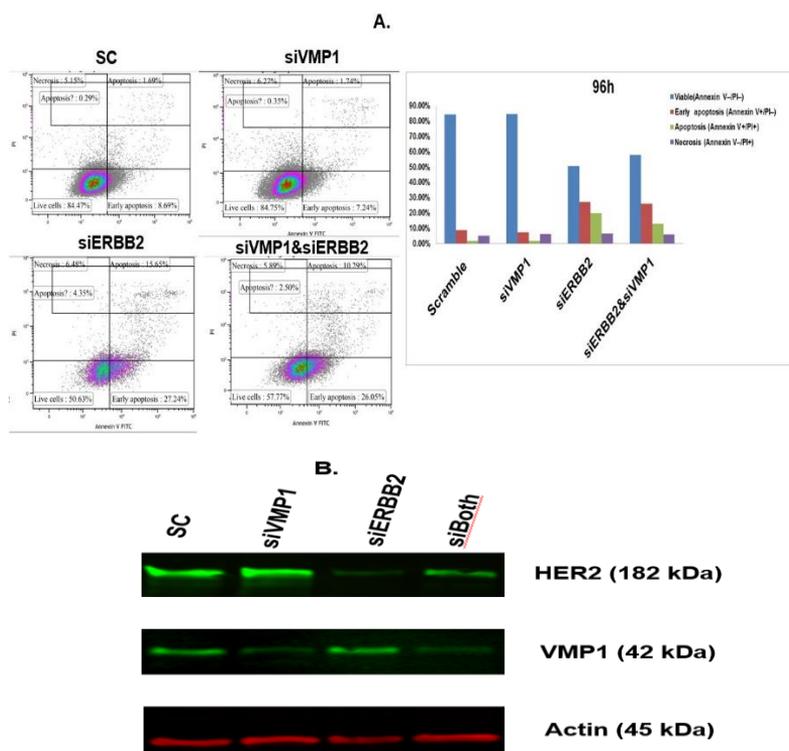


Figure 29. VMP1 knockdown did not induce apoptosis.

BT474 cells were treated with SC, siVMP1, siERBB2 and siVMP1+siERBB2; 96 hours after adding siRNAs, cells were lysed to evaluate apoptosis status. siERBB2 was used as positive control and SC as negative control, and all experiments were performed in triplicate and repeated two times. A. shows the percentage of viable cells, early apoptosis, apoptosis, and necrotic cells targeted with SC, siVMP1, siERBB2 and siERBB2. B. represents confirmation of KD in this experiment.

4.4.6 Effect of VMP1 knockdown on proteins associated with cell adhesion in BT-474 cells

VMP1 has a role in cell adhesion and formation of tight junctions (Sauermaun et al., 2008). Assessment of VMP1 mRNA within breast cancer cohorts demonstrated higher expression of VMP1 in tumors with metastasis and node positive tumors and an association with shorter distant recurrence time of patients. All the above-mentioned information led us to investigate the role silencing of VMP1 has on expression of the tight junction proteins, ZO-1 and E-CAD. ZO-1 is a protein that comprises part of the framework of tight junction transmembrane proteins (Bhat et al., 2018). E-CAD is involved in formation of adherent junctions (Hartsock & Nelson, 2008).

4.4.6.1 VMP1 silencing does not affect ZO-1 expression

To test the effect of VMP1 silencing on metastasis and the structure of tight junctions, lysates from BT-474 cells treated with siRNAs were probed with VMP1, HER2 and ZO-1 primary antibodies, and signals normalized to actin were compared. Silencing VMP1 and ERBB2, alone or in combination, had no effect on expression of the tight junction protein, ZO-1 (Figure 30).

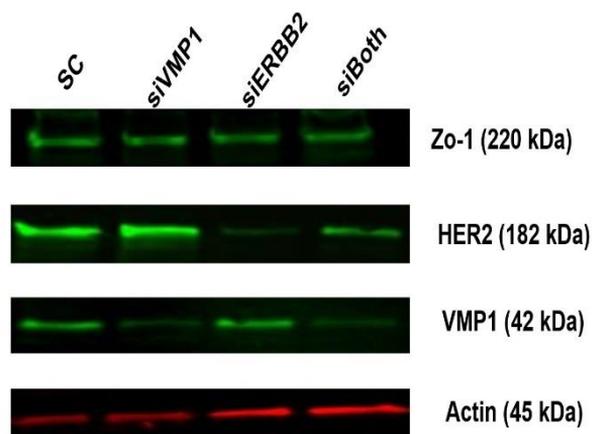


Figure 30. Silencing VMP1 did not affect ZO-1 expression. Effect of silencing VMP1 and ERBB2, alone or together, on expression of ZO-1 protein was analyzed with Western blot 72 hours after transfecting BT474 cells with siRNAs and compared with scramble siRNA.

4.4.6.2 Silencing VMP1 did not have effect on expression of E-CAD

During proliferation assays in BT474 cells, the VMP1 knockdown caused cells to change shape, reminiscent of changes in cells going through the EMT. Nevertheless, VMP1 silencing had no effect on E-CAD expression (unlike ERBB2 silencing, which lowered expression of E-CAD protein (Figure 31). This experiment was done only one time and should be repeated with other EMT markers such is Vimentin, N-Cadherin and SNAILS.

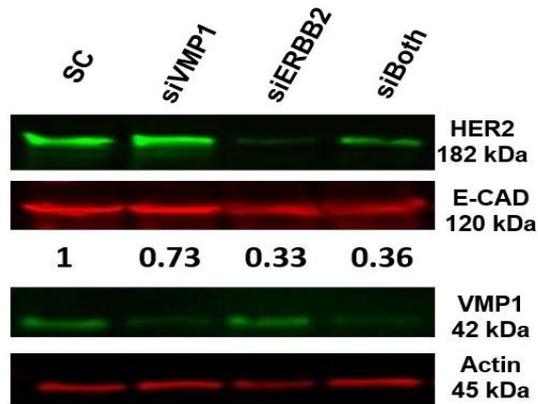


Figure 31. VMP1 silencing does not affect E-CAD expression. Effect of silencing VMP1, ERBB2 and both on expression of E-CAD protein was analyzed with Western blot 72 hours after transfecting BT474 cells with siRNAs.

Taking the data from this section as a whole, VMP1 mRNA was highly expressed in breast cancer cell lines and ER-positive and HER2 positive tumors. The silencing of VMP1 affected neither HER2 expression nor BT474 cell proliferation. Other experiments like apoptosis assays and the effect of silencing VMP1 on ZO-1 and E-CAD should be repeated.

5 Technical hurdles and considerations

During the course of the project, multiple methods and techniques were established and standardized at the lab. For each experiment, there were technical aspects that must be considered and many hurdles along the way. Here I will discuss some of the technical difficulties that arose in this project.

5.1 Confirmation of fusion genes

For confirming the existence of fusion genes identified through our bioinformatic pipelines and *in silico* analysis, primers that spanned junction sites of gene partners of fusion genes were designed. Using this method, we could confirm (several) fusion genes previously identified, but we had difficulty with confirming the VGLL4:SH3BP5 and ESR1:CCDC170 fusion genes. Designing new primer sets and changing the melting temperature of the primers did not solve the problem. We sought to solve the problem by direct extraction of amplified products from the gel and perform Sanger sequencing on the isolated product, however, we could not perform Sanger sequencing due to extremely low yield of extracted amplified products from the gel and therefore the presence of these fusion genes was not confirmed.

5.2 Immunohistochemistry

One of the main technical hurdles of this thesis was the staining of tumor and normal breast tissues with VMP1 antibody in a tissue microarray (TMA). Staining of tissues was performed with Anti-TMEM49 (VMP1)/ (2790506) antibody C-terminal (Abcam) with the 1/100 recommended concentration, as per the manufacturer. Unfortunately, the antibody did not show specific staining, stained all parts of tissues, and gave very high background throughout the tissue. This made it hard to score the TMA slides. Increasing and decreasing the concentration of the antibody and the incubation time did not change the results. Using other antibodies from Abcam: ab116006/Anti-TMEM49 antibody and ab203684/Anti-TMEM49 antibody-C-terminal did not change our results and gave unspecific staining.

Abcam: ab116006/Anti-TMEM49 antibody has worked in a study for staining of human colorectal cancer tissues and matched adjacent non-cancerous tissues (X. Z. Guo et al., 2015). They used an immunohistochemical kit for staining of tissues. In our study we used the

same antibody which X. Z. Guo et al used but got inconsistent observation with their study. This might be due to using different staining methods and distinct types of tumors.

Tissue microarray (TMA) is a high throughput method for screening of protein expression patterns within tissue samples from large patient cohorts. TMA contains only limited amount of tissue. Due to this tissue heterogeneity can be an issue and several samples may be needed from the same specimen. For tumor tissues it is highly recommended to use two to four cores from each specimen (Kampf, Olsson, Ryberg, Sjöstedt, & Pontén, 2012). Breast tumors tissues are complex and composed of several different cell types, structures and extra-cellular matrix. They contain variable amount of fat, blood vessels and fibrous. The heterogeneity of breast tumors and normal tissues will affect the pressure needed for punching and collecting separate cores for preparing of TMA slides and affinity of primary antibodies. VMP1 is a transmembrane protein and localized to endoplasmic reticulum and nucleoli("The Human Protein Atlas (<https://www.proteinatlas.org/>),"). Localization of VMP1 protein in both nucleoli, cytoplasm and cells membranes might be reason for unspecific staining of our TMA breast tumor and normal samples with VMP1 antibody.

5.3 Slow growing cell lines

One of the technical hurdles of cell-based experiments was the relatively long doubling time of the BT474 and MDA-MB-361 lines, which affected the timeframe and setup of many experiments (doubling time of BT474 and MDA-MB-361 cells is 3.5 and 4 days, respectively). Using siRNAs in BT474 and MDA-MB-361 cells to knockdown of VMP1 and ERBB2 genes was difficult because the effect of siRNAs lasts for 96 hours, at most.

5.4 Transfection of siRNAs

For knock down of VMP1 and ERBB2 in BT474 and MDA-MB-361 we initially followed the protocol from Ambion (MAN0007836) and performed traditional transfection where cells are seeded into the cell culture plate 24 hours before transfection. Using this method, the VMP1 knock down efficiency in BT474 cells was 63% and 23% in MDA-MB-361 72 hours after adding siRNAs. Using the same method for ERBB2 knock down the efficiency for BT474 and MDA-MB-361 cells were 60% and 70%, respectively. It is likely that their morphology might affect the siRNA transfection efficiency 24 hours after cell seeding. For this reason, we changed our protocol to "reverse transfection" where the siRNAs are to cells at the time of seeding. This method attained a

92% knockdown for VMP1 in BT474 cells and 67% knockdown in MDA-MB-361 cells. With this method, ERBB2 knockdown efficiency was 92% in BT474 cells and 82% in MDA-MB-361 cells.

Knocking down of VMP1 gene using small interfering RNA (siRNA) is a temporary method for silencing it. Due to the relatively long doubling time of most of HER2 positive breast cancer cell lines, for studying VMP1's role within them it would be efficient to use more stable knock down method. Using of another knock down method such as short hairpin RNA (shRNA) might give longer and more stable VMP1 gene silencing. siRNA and shRNA gene knock down methods traditionally involved interfering with mRNAs or non-coding RNAs have natively produced by cells and they do not affect and involve host DNA. They are temporarily decrease and stop the expression of targeted genes. The cells may survive a knockdown event and can recover. This leads to expression of the gene as before. In newer knockdown techniques such as CRISPR genome editing using of dCas9 protein, which is a mutant, enzymatically dead form of the CRISPR-associated protein Cas9 interacting with host DNA instead of interfering with host RNA (Tian et al., 2019). Usage of CRISPR dCas9 technology of for knockdown of VMP1 gene might be more efficient than traditionally knock down methods such as siRNAs and shRNAs.

5.5 Alternative method for finding link between VMP1 and different proteins

VMP1 has diverse biological and pathological roles. As several lines of evidence from studies of VMP1 show its interaction with various proteins. To investigate VMP1's role in breast cancer particularly HER2 positive tumors, it would be interesting to find all the proteins that cooperating with VMP1. To identify the key target genes cooperating and interacting with VMP1 gene within breast cancer cell lines, instead of picking up individual genes/proteins in VMP1 and VMP1/ERBB2 knock down cells it would be useful to use high-throughput sequencing such as RNA sequencing (RNA-seq). Using of this technique would provide comprehensive information about differential gene expression analysis within VMP1 and VMP1/ERBB2 knock down cells.

6 Discussion

Given the role of fusion genes in carcinogenesis, we speculated genes recurrently found as might reflect their role in breast cancer development. Vacuole membrane protein 1 (VMP1) was found through this approach. Screening of two Icelandic breast cohorts and confirming the results in two large, publicly available breast cancer cohorts, we identified VMP1 as a gene with a potential role in the development of breast tumors, particularly the HER2 positive subtype. Due to location of MIR21 and considerable sequence overlap between it and VMP1, we postulated expression from the two genes could affect one another. Although hsa-miR-21-5p is a well-known oncomir in breast cancer, its “sibling” hsa-miR-21-3p is hardly studied at all. Thus, we also found that hsa-miR-21-3p, which is transcribed from its own promoter within intron 10 of VMP1, is a potential marker in development of breast tumors. Data from these two studies showed that screening of fusion genes can be a potential approach for finding of novel breast cancer genes associated with progression.

6.1 Identification of breast cancer genes

Our data suggest screening for fusion genes, in both cell lines and tumors, can identify genes that might play a role in cancer development. In our study, we screened for fusion genes in both breast cancer cell lines and breast tumors and, after applying filtering criteria, five fusion genes were identified as linked to breast cancer. Fusion genes are not cancer-specific features but are also found in non-cancerous tissues (Anca Botezatu, 2016; Babiceanu et al., 2016; Yan et al., 2016). Since appearance of stromal and normal cells and microenvironment components within breast tumor cells are unavoidable (Januškevičienė & Petrikaitė, 2019), noncancerous fusions must be carefully filtered from analysis.

Cancer cell lines tend to be aggressive and most lack stromal and normal cell contamination (Mirabelli, Coppola, & Salvatore, 2019), which makes rearrangement detection easier (Inaki et al., 2011). One of the novelties of our screening method was comparison of fusion genes from breast cancer cell lines and tumors, to filter out normal cell contamination within tumors and increase the feasibility of detecting genes associated with breast cancer progression. Most gene-fusion studies focus on chimeric fusion proteins

(Asmann et al., 2012; Nik-Zainal, 2016; Yoshihara et al., 2015), because they often generate targetable proteins like BCR-ABL1 (Quintás-Cardama & Cortes, 2009; Quintás-Cardama, Kantarjian, & Cortes, 2009). In contrast, our screening method did not focus on the fusion genes themselves but on the single genes that constitute the fusion.

In our first screening step, 15 fusion genes were identified that were found in both cell lines and tumors. Of the breast cancer cell lines included in our model, MCF7 had the highest number of fusion genes (five out of 15), followed by UACC893 cells (two out of 15). According to the viewpoint of Paul A W Edwards and Karen D Howarth (Edwards & Howarth, 2012), who reviewed papers from Chinnaiyan AM et al. (Robinson et al., 2011) and Milosavljevic A et al. (Hampton et al., 2009; Hampton et al., 2011), MCF7 had the highest number of fusion genes. Thus, the data from these studies support our model and MCF7 may be a well-established model cell line for studying and discovering new fusion genes. The ITGB6:RBMS1 fusion was a novel fusion gene reported and confirmed for the first time by us in the UACC893 cell line.

There were differences between the 15 fusion genes identified in breast cancer cell lines and the tumors. Interestingly a third of 15 fusion genes had similar break points in both cell lines and tumors, while two thirds had different break points in cell lines and tumors. This diversity of fusion genes between breast cancer cell lines and tumors may be linked with tumor heterogeneity or an effect of the tumor microenvironment. Also, some of the fusion genes had open reading frames (ORFs), while other fusion genes did not; and the number of fusion genes with ORFs was different between cell lines and tumors.

Our filtering criteria included showing an identical breakpoint in breast tumors and cell lines, being recurrent in tumors, not be located within an amplicon carrying a known oncogene unless it was part of the fusion and having a function supportive of tumorigenesis (available through publications). Five fusion genes, or ten individual genes, passed these criteria.

DNA amplifications are found frequently in breast tumors, where breakage and rejoining of chromosomes occurs (Shiu, Natrajan, Geyer, Ashworth, & Reis-Filho, 2010). As a result, fusion genes are often found in amplified regions. Genes amplified in concordance with an increase in expression have been implicated in cancer (Ohshima et al., 2017). Therefore, the final filtering

criterion was the correlation between expression of DNA and mRNA, which led to the identification of VMP1 and RPS6KB1 genes.

Tandem duplication of the adjacent genes, VMP1 and RPS6KB1, resulted in formation of the RPS6KB1:VMP1 fusion gene. This fusion gene was the second most frequent recurrent fusion gene in our analysis (5/1724).

According to Inaki, K et al.'s study in 70 breast cancer patients from Singapore, the RPS6KB1:VMP1 fusion gene was observed in 22 tumors of their study group. They identified multiple types of fusions between the two genes, but most did not have an ORF. The most frequent fusion types from their study, detected in 10 cases, was in E1 of RPS6KB1 and E8 of VMP1, followed by nine cases which detected fusion of in E1 of RPS6KB1 and E11 of VMP1 (Inaki et al., 2011) .

The number of RPS6KB1:VMP1 fusions (22/70) in the Inaki, K et al. study was inconsistent with the numbers in our model (5/1724). In our model this fusion detected through *in silico* analysis from paired-end RNA sequencing data. In contrast, they analyzed tumors with probes specific for the RPS6KB1:VMP1 fusion, which might identify this fusion in tumors of patients of European decent, via RT-PCR. The RPS6KB1:VMP1 fusion was expressed at low levels in normal breast tissue and the chimeric protein did not contain any functional domain implicating it in development of breast cancer (Inaki et al., 2011; Veeraraghavan, Ma, Hu, & Wang, 2016). The association of RPS6KB1 with HER2 positivity and a worse outcome was shown (Pérez-Tenorio et al., 2011). RPS6KB1 did not associate with overall survival in the TCGA cohort (Appendix figure 4) whereas VMP1 did (Figure 9). Taken together, these data support that Vacuole membrane protein 1 (VMP1) is the strongest candidate for further follow-up.

6.2 Vacuole membrane protein 1 (VMP1)

This chapter discusses the results from an association analyses of VMP1 mRNA with clinical and pathological characteristics, as well as survival analyses. VMP1 mRNA was expressed at higher levels in breast tumors than matched normal breast tissues and higher in HER2 positive tumors than HER2 negative tumors. Elevated levels of VMP1 mRNA were a marker of worse survival in breast cancer patients, particularly those with HER2 positive tumors.

6.2.1 Biology

Vacuole membrane protein 1 (VMP1) is a multi-spanning transmembrane

protein localized in the endoplasmic reticulum (ER), where it forms micro domains (Calvo-Garrido, Carilla-Latorre, Lázaro-Diéguez, Egea, & Escalante, 2008). VMP1's localization in the ER is in close contact with mitochondria (Tábara et al., 2018) and was originally identified as a protein associated with pancreatitis since its overexpression in rats with acute pancreatitis induces formation of vacuoles and cell death (Duseti et al., 2002). VMP1 has a role as an inducer of autophagosome formation (Vaccaro, Ropolo, Grasso, & Iovanna, 2008) and is one of the main components in the formation of tight junctions. VMP1 interacts with the Zonula Occludens and is involved in cell adhesion, invasion, and metastasis (Sauermaun et al., 2008). It is essential for survival during the early embryonic period in zebrafish and mice (Morishita et al., 2019); and in organs such as intestine, liver, and visceral endoderm, VMP1 is important molecule for release of lipoprotein from ER membrane to the lumen (Morishita et al., 2019).

Autophagy is a process with multiple steps and with many participating proteins (Li, He, & Ma, 2020). The initial steps include assembly of membranes that form the autophagosome, a process that VMP1 is involved in (Molejon, Ropolo, Re, Boggio, & Vaccaro, 2013). Beclin 1 is the main regulator of the initiation of the autophagic process (Al-Bari, 2020). The Atg domain in the C-terminus of VMP1 interacts with Beclin 1. BH3 domain of apoptotic proteins like Bcl-2 or Bcl-XL also have the ability of interaction with the BH3 domain of Beclin1. Bcl-2 and Bcl-X diminish Beclin 1's pro-autophagic activity but in contrast Beclin 1 cannot abrogate the pro-apoptotic activity of Bcl proteins (Kang, Zeh, Lotze, & Tang, 2011). Autophagy has a dual role as suppressor and promoter in cancer (Russo & Russo, 2018). Autophagy levels fluctuate during carcinogenesis: in initial stages autophagy inhibits tumor formation but in late stages autophagy promotes tumor formation (Maes, Rubio, Garg, & Agostinis, 2013). Uncontrollable growth of cancer cells leads to a demand of nutrients and energy in the tumor's microenvironment; and autophagy is a well-established survival mechanism for tumor cells in this condition. Upregulation of PI3K/Akt signaling and loss of tumor suppressor PTEN result in inhibition of autophagy (Singh et al., 2018). Due to the role of VMP1 in this pathway, further studies are necessary to investigate the connection of hyperactivation of the PI3K/AKT pathway and elevated VMP1 expression in the HER2 pathway for tumor development.

6.2.2 VMP1's role in breast cancer

A few studies have evaluated VMP1's role in breast tumors. Sauermaun et al. quantified VMP1 mRNA levels in invasive ductal carcinoma patients (n=45)

and found negative correlation between increasing grade of breast tumors and VMP1 mRNA levels (Sauermann et al., 2008). In our study, VMP1 did not associate with tumor histograde. There are a few potential explanations for the discrepancy. The cohort that Sauermann et al. used in their study was limited to only invasive ductal carcinoma (IDC) whereas, in our study, the tumors were a mixture of ductal, lobular, and other histopathological subtypes. The number of patients they included in their study was small (n=45). Their analysis of only IDC tumors did not confirm our result, i.e., VMP1 mRNA expression did not correlate with grade.

In a recent study of 94 breast cancer tissues and 54 matched adjacent non-cancerous tissues, VMP1 protein was quantified by immunohistochemistry, showing strong cytoplasmic staining of VMP1 protein in non-cancerous tissues than in high stage breast tumors which had weak or no positive cytoplasmic staining. According to their results VMP1 was higher in ductal carcinoma in situ (DCIS) than in invasive ductal carcinoma (IDC) (81.3% vs 56.3%; $\chi^2=4.655$, $P=0.031$), and patients who were negative for expression of VMP1 protein had shorter disease free survival and worse prognosis than patients who had tumors that were positive for expression of VMP1 protein (Sun et al., 2019). Our findings were not in accord with their findings. The reasons for the discrepancy may relate to their study group in which majority of tumors were DCIS (78.94%) whereas all tumors in the four cohorts in our analysis were primary breast tumors. Their cohort had fewer patient samples than ours, which might also have caused the discrepancy. Other reasons might be due to using different statistic methods for analyzing data within two studies: they used chi-square test, but we used the T-test and Anova. They used immunohistochemistry for quantification of VMP1 protein, but our immunohistochemistry did not work so we used VMP1 mRNA in lieu of VMP1 protein. All these factors might have caused the discrepancy between our results and theirs.

6.2.3 VMP1's role in other cancers

VMP1 has been suggested as having tumor suppressive properties in several types of tumors. In hepatocellular carcinoma (HCC) reduced VMP1 expression suppressed metastasis and VMP1 was found as a functional target of miR-210. Expression of miR-210 and VMP1 was inversely correlated; and downregulation of VMP1 by miR-210 mediated induction of hypoxia and related metastasis of tumor cells (Ying et al., 2011). Another study in ovarian cancer cells, showed suppression of invasion and migration capacity of tumor cells after suppression of VMP1 expression by HIF-1 α /miR-

210 through increased expression of VHL (T. Liu et al., 2014). Direct down regulation of VMP1 by miR-210 was shown in colorectal cancer as well (Qu et al., 2014). In colorectal cancer, VMP1 was shown to be a regulator of apoptosis and autophagy (Qian et al., 2014), its expression was higher in adjacent non-cancerous colorectal tissues than cancerous tissues, and downregulation of VMP1 correlated with shorter survival time of patients (X. Z. Guo et al., 2015). This conflicts with our findings on the effect of VMP1 in breast cancer. Guo et al. showed downregulation of VMP1 as a marker of poor prognosis in HCC (L. Guo, Yang, Fan, Chen, & Wu, 2012).

These discrepancies regarding VMP1's role cancer could be due to tumor type and the role of VMP1 in autophagy. Overexpression of VMP1 triggers formation of autophagosomes and autophagy levels fluctuate during development of tumors, in initial stages it inhibits tumor formation but in late stages autophagy promotes tumor formation (Maes et al., 2013). In cancer, the dual nature of autophagy genes as both oncogene and tumor suppressor depends on context (Singh et al., 2018), reflecting dissimilar roles during breast tumorigenesis (Céline Grandvallet, 2020).

6.2.4 VMP1 and HER2

VMP1 expression levels were higher in HER2 positive tumors than negative tumors within all four breast cohorts in this study (Figure 12). The 17q23 locus, where VMP1 resides, is amplified in 20% of ERBB2 amplified tumors (Jonsson et al., 2010; Staaf et al., 2010). In a study of pancreatic cancer cells, KRAS^{G12D} was an inducer of the AKT1-GLI3-p300 signaling pathway. Induction of the AKT1-GLI3-p300 pathway leads to upregulation of VMP1 and subsequent initiation of autophagy (Lo Ré et al., 2012). In breast cancer, gain-of function-mutations in ERBB2 and EGFR lead to hyperactivation of the PI3K/AKT pathway (Carmona et al., 2016), which may result in upregulation of VMP1, one of the downstream molecules of this pathway.

During initiation of autophagy and autophagosome formation, the C-terminus of VMP1 interacts with Beclin 1 (Molejon, Ropolo, Re, et al., 2013). Autophagy has a multi-faceted role in tumorigenesis and metastasis and its levels fluctuate during different stages of tumor development (Dower, Wills, Frisch, & Wang, 2018; Maes et al., 2013). On one hand, binding of HER2 to BECN1 in HER2 positive breast cancer cell lines were shown to inhibit autophagy (Han et al., 2013; Vega-Rubín-de-Celis et al., 2018). HER2 does not phosphorylate BECN1 and alters autophagy through a mechanism independent of BECN1 phosphorylation (Vega-Rubín-de-Celis, 2019). On the other hand interaction between VMP1-AtgD and BECN1-BH3 domains leads

to localization of Class III PI3K activity on the autophagosome formation site and induces autophagy (Molejon, Ropolo, Re, et al., 2013). VMP1 and HER2 interact with a common protein, BECN1, whose interaction with HER2 inhibits autophagy, yet the VMP1/BECN1 interaction induces autophagy. Interaction of VMP1 with BECN1, in various stages of tumor development in HER2 positive cells, activates autophagy and that may lead to metastasis of HER2 positive cells, an issue that must be addressed in functional cell-based assays.

Breast cancer patients with HER2 positive tumors with high VMP1 expression levels within the METABRIC cohort (n= 1220) had shorter BCSS and DRFS survival in comparison with HER2 negative ones (Figure 17). In the clinic, in the long run, almost 50-70% of HER2 positive breast cancer patients do not benefit from treatment due to de-novo and acquired resistance that leads to relapse (Hudis, 2007; H. Jiang & Rugo, 2015; Pohlmann et al., 2009). Acquired resistance to anti-HER2 therapy, particularly trastuzumab, may arise through many mechanisms: Activation of compensatory pathways; mutation of the HER2 receptor; gene amplification and increased expression (Pohlmann et al., 2009). Various studies show activation of autophagy and its crosstalk with apoptosis causes resistance to lapatinib and trastuzumab in HER2 positive breast tumors (Janser et al., 2019; Mele et al., 2020; Zambrano & Yeh, 2016).

Due to the roles of VMP1 in the AKT1-GLI3-p300 signaling pathway and initiation of autophagy, VMP1 overexpresses in HER2 positive tumors (as we observed in METABRIC/HER2 positive tumors; Figure 17B) may induce metastasis and relapse of HER2 positive tumors. VMP1 also may play a role in resistance particularly trastuzumab through its role in autophagy. However, more investigation is needed regarding VMP1' s role in survival and therapy resistance of HER2 positive breast cancer cells.

6.2.5 Discrepant role of VMP1 within cohorts

Patients with elevated levels of VMP1 mRNA in cohort 1 and METABRIC had shorter BCSS and DRFS but no such effect was observed in cohort 2 and TCGA. One explanation for the discrepant role of high VMP1 in the four cohorts in this study may relate to the extended period of tumor collection. Patients in cohort 1 were diagnosed between 1987 and 2003, in cohort 2, 2003 to 2007, TCGA 1987 to 2013, and METABRIC 1980 to 2005. Different time of diagnosis leads to different drug treatments.

Another reason for the discrepancy may relate to a patient's changing treatment protocols, over time, and prescription of new drugs such as trastuzumab. Trastuzumab was a breakthrough in treatment of HER2 positive breast cancer when it became available around the turn of the century. (It got FDA approval on September 25, 1998). At the beginning it was available only to women with HER2 positive metastatic breast cancer (Slamon et al., 2001) and was not until Nov 16, 2006 that it was approved for adjuvant treatment of primary HER2 tumors (Joy & Mackey, 2006). Notably, the breast cancer patients in the METABRIC cohort did not receive trastuzumab (Curtis et al., 2012). Patients in TCGA were diagnosed over a period of 26 years (1987 to 2013) and they, most likely, received different therapy regimens. None of the HER2 positive patients in cohort 1 received trastuzumab, because all were diagnosed before approval of the drug in Iceland. HER2 positive patients in cohort 2 received trastuzumab because they were all diagnosed after approval of trastuzumab in Iceland.

6.3 Hsa-miR-21-3p

This part of the study describes analysis of hsa-miR-21-3p expression in cohorts 1 and 2 and the correlation with clinical and pathological parameters as well as survival analyses among the four cohorts. Hsa-miR-21-3p was higher in breast tumors than matched normal breast tissues and its expression was associated with HER2 positivity, ER negativity, advanced tumor stage, large tumor size, high histograde, and positive node status. High hsa-miR-21-3p levels were a marker of worse survival in breast cancer patients.

Hsa-miR-21-5p is one of the most extensively studied microRNAs in cancer including breast cancer. In BC, high expression levels are associated with advanced-stage tumors, nodal status, and poor prognosis (Y. H. Feng & Tsao, 2016; O'Bryan, Dong, Mathis, & Alahari, 2017). The number of publications examining the role of hsa-miR-21-3p in breast cancer are just a handful compared to publications on its frequently cancer-associated -5p counterpart, so this project focused on hsa-miR-21-3p. To explore the role of hsa-miR-21-3p in breast cancer, we quantified levels of both hsa-miR-21-5p and hsa-miR-21-3p in breast cancer cohorts. Data regarding hsa-miR-21-5p are not shown in the Result section but can be found in the Appendix (Tables 15, 16,17, and 18).

Based on our results, hsa-miR-21-5p was expressed at higher levels in HER2 positive tumors in cohort 2 ($p=0.002$); and in TCGA it was higher in

patients with age below 50 ($p=0.008$). In METABRIC/EGA, its expression was higher in PGR negative ($p= 0.003$), and HER2 positive ($p= 1.01 \times e^{-14}$) tumors. hsa-miR-21-5p did not associate with survival for any of the cohorts.

6.3.1 The role of hsa-miR-21-3p in breast cancer and other cancer types

In a study of triple-negative breast tumors from TCGA, hsa-miR-21-3p was identified as an independent risk factor for overall survival (OS) (X. Wu, Ding, & Lin, 2020); however, we did not observe an association between hsa-miR-21-3p and survival in TNBC in TCGA ($n=46$) and METABRIC/EGA ($n=197$). Notably, the Wu et al. study used RNA sequencing data for their analysis, whereas we used VMP1 measurements from the METABRIC and TCGA which were quantified with microarray technology. In miRBase, there are other isomiRs for the canonical sequence of hsa-miR-21-3p (Harbeck et al., 2019). Different features of hsa-miR-21-3p isomiRs were found in colon cancer (Jiao et al., 2017). Since microarray probes are long, they likely pick up multiple isomiRs.

In another study, hsa-miR-21-3p was identified as a miRNA that is upregulated in triple negative tumors; again, a result that does not agree with our findings (Ouyang et al., 2014). Nevertheless, in their study, when subjects were divided into two groups based on hsa-miR-21-3p DNA content (gain and no gain), hsa-miR-21-3p expression was higher in the gain group, which agrees with our study data. Based on their functional analysis overexpression of hsa-miR-21-3p in KPL-4 and MCF-7 cells lines resulted in increased proliferation measured with Ki67. In addition, they demonstrated that overexpression of hsa-miR-21-3p resulted in an increase of phosphorylated AKT protein in BC cell lines, JIMT-1 and KPL-4 cells (Aure et al., 2013).

In a study seeking to develop a method to identify noninvasive biomarkers, Xiaokang Yu, et al. used plasma as an approach. First, RNA sequencing data (from 409 breast tumor tissues and 87 healthy controls from TCGA) was used at the discovery stage. Next, serum from 113 breast cancer patients in initial stages and 47 healthy controls were used in a validation stage. Hsa-miR-21-3p was among 11 upregulated microRNAs within breast tumor tissues and serums. Hsa-miR-21-3p expression levels were significantly higher in plasma of early-stage breast cancer patients than that of healthy controls. Taken together, their data implicates hsa-miR21-3p as a potential biomarker for early detection of breast cancer (X. Yu et al., 2018).

In a study using serum and tissue from 20 non-small cell lung cancer patients (NSCLC) and adjacent normal tissues (20 benign lung disease and 20 healthy volunteers), hsa-miR-21-3p was expressed at higher levels in the serum of lung cancer patients than the two other groups and the adjacent normal tissues (M. Jiang et al., 2013). Despite the use of different tumor types in their study and ours, our study showed the expression levels of hsa-miR-21-3p in both cohort 2 and TCGA were higher in tumors than normal tissues, as did theirs in NSCLC.

A study that Lu et al. performed in laryngeal carcinoma tissue and para-cancerous tissue samples, hsa-miR-21-3p was demonstrated as upregulated in cancerous tissues, again in line with our data (Lu et al., 2014). L1 cell adhesion molecule L1CAM has roles in cell motility, invasion, metastasis formation and chemoresistance. K. Doberstein et al. found expression of L1CAM upregulated with hsa-miR-21-3p in cancer cell lines in renal cell carcinoma (RCC), endometrial carcinoma and ovarian carcinoma origins and observed strong positive correlation between hsa-miR-21-3p and L1CAM in patients with RCC ($r=0.76$, $P=0.02$), (Doberstein et al., 2014). Their data support the idea that hsa-miR-21-3p is an oncogene. R.C. Pink et al. measured miRNA levels in ovarian cancer cell lines and their cisplatin resistance clone, and found the resistant clones expressed more hsa-miR-21-3p. As a potential target for hsa-miR-21-3p, they suggested the neuron navigator 3 (NAV3) gene, (Pink et al., 2015), a known tumor suppressor (Carlsson et al., 2013; Karenko et al., 2005; Ranki, Väkevä, Sipilä, & Krohn, 2011). This suggests the microRNA could have a role in drug resistance and might be a drug target.

In a study in esophageal squamous cell carcinoma (ESCC), tumor and para tumor tissues showed differential expression of hsa-miR-21-3p, in agreement with our findings. Hsa-miR-21-3p also was shown as inducer of proliferation, migration, and invasion in ESCC (Z. Gao et al., 2019). Inhibition of proliferation, invasion, tumor growth and promotion of apoptosis with increased hsa-miR-21-3p was reported in ovarian cancer cells and hepatocellular carcinoma, however in colorectal cancer hsa-miR-21-3p upregulation promoted cell adhesion and invasion (Báez-Vega et al., 2016) (Hou et al., 2018; Lo et al., 2013). Taken together, these data suggest hsa-miR-21-3p has a dual role, both as both a tumor suppressor and oncogene, depending on tumor type. This might be due to different cancer types and targeting both tumor suppressor and oncogenes through hsa-miR-21-3p.

6.3.2 The discrepant role of hsa-miR-21-3p within cohorts

Patients with elevated levels of hsa-miR-21-3p in cohort 1 and METABRIC/EGA had a significantly shorter DFS, an effect not seen in cohort 2 and TCGA. Elevated levels of hsa-miR-21-3p associated with shorter BCSS only in METABRIC/EGA. Hsa-miR-21-3p Pearson correlation value with ERBB2 mRNA within TCGA and METABRIC/EGA were ($r=0.14$, $p=0.02$) and ($r=0.13$, $p=5.18e^{-6}$), respectively. These coefficients imply there is no relationship between ERBB2 mRNA and miR21-3p.

Due to well established oncogenic role of HER2 in breast tumors, hsa-miR-21-3p and ERBB2 mRNA correlation and high expression of hsa-miR-21-3p in HER2 positive tumors, there was a possibility that HER2 confounded the effect of high hsa-miR-21-3p on survival. For this reason, whether HER2 positive status affects the impact of high hsa-miR-21-3p expression on DFS and BCSS was checked with Cox regression analysis. The effect of high hsa-miR-21-3p expression on survival was attenuated with HER2 in cohort 1 and METABRIC/EGA. Over the years, numerous studies of breast tumors showed that ER, HER2, Grade, and node status were linked to patient survival. For this reason, the effect of above-mentioned factors was checked with Cox regression analysis, asking whether high expression of hsa-miR-21-3p affects survival. Although the effect of elevated levels of hsa-miR-21-3p on survival was confounded by ER, HER2, node and grade, after adjusting for those features, the p-value remained significant (Table 17). The largest confounding effect was high grade, but the effect of hsa-miR-21-3p expression stayed significant after adjusting to grade (HR=1.25, CI: 1.02 - 1.52), implicating hsa-miR-21-3p as an independent marker associated with shorter survival. Cox regression analyses also performed checked whether ER, HER2, node and grade confounded the effect of hsa-miR-21-3p on DFS in cohort 1 and METABRIC/EGA. Grade and HER2 were strongest confounders but the effect of hsa-miR-21-3p on DFS stayed significant after adjusting to grade and HER2, in both cohort 1 and METABRIC/EGA. Due to the high correlation values between hsa-miR-21-3p and hsa-miR-21-5p, RPS6KB1, and PPM1D, and their reported oncogenic effects, Cox regression analyses was performed to adjust for potentially confounding effects of high RPS6KB1, PPM1D and hsa-miR-21-5p on hsa-miR-21-3p in METABRIC/EGA cohort. Here, the effect of high hsa-miR-21-3p on BCSS was not confounded by hsa-miR-21-5p, RPS6KB1 and PPM1D genes.

As emphasized in subsection 5.2.5, the patients in cohort 1 (1987–2003), cohort 2 (2003–2007), TCGA (1987–2013) and METABRIC/EGA (1980-

2005), all got different therapy regimes because they had different diagnoses. This may explain the discrepant effect of high hsa-miR-21-3p on survival within cohorts.

6.3.3 Correlation of genes located at chromosome 17q23 locus

Many genes reside at this locus, but only RPS6KB1, MIR21 (Haverty et al., 2008), PPM1D (Natrajan et al., 2009) and VMP1 (Amirfallah et al., 2019) have been shown to be oncogenes. VMP1 and miR-21 have their own promoters but overlap for maximum 4kb of genome, and sharing 1kb of RNA sequence (Z. Wang, 2013) (Figure 5).

In TCGA, cohort 1, and cohort 2, the probes for detecting VMP1 mRNA hybridize to the C-terminus, so they could detect pri-miR-21. For this reason, another probe, spanning exons 2-3 (encoding the N-terminus) of the VMP1 gene was used. Correlation of another probe located in N-terminus of VMP1 mRNA in cohort 1 with the probe located on C-terminus was high ($r = 0.85$, $p < 0.001$) and none of the probes for the C-terminus and N-terminus correlated with hsa-miR-21-3p and hsa-miR-21-5p quantities in cohort 1 ($p > 0.05$). In cohort 1, the VMP1 and hsa-miR-21-3p Pearson correlation value was ($r=0.08$, $p= 0.32$). In cohort 2 it was ($r=0.03$, $p= 0.62$); in TCGA was ($r=0.42$, $p= 2.26e^{-49}$); and in METABRIC, it was ($r=0.57$, $p<2.2e^{-16}$). The hsa-miR-21-3p and VMP1 signals were measured by different methods within the five cohorts. Quantification of hsa-miR-21-3p in cohort 1 and cohort 2 was performed with qRT-PCR, using a probe that specifically detects the reference hsa-miR-21-3p sequence. In both TCGA and METABRIC/EGA, quantification of hsa-miR-21-3p was performed with microarrays. The average probe length in this technology is 40 to 60 nucleotides (www.agilent.com), which is three times longer than the length of mature microRNAs. Therefore, the microarray probes may catch other isoforms of hsa-miR-21-3p, VMP-miR-21 and pri-miR-21. VMP1 levels, like those of hsa-miR-21-3p, were measured with qRT-PCR in cohort 1 and cohort 2. The Taqman probe for measuring VMP1 levels in cohort 1 and cohort 2 spans exons 10 and 11. In TCGA and METABRIC cohorts, VMP1 quantification was performed using microarray technology. Two microarray probes in TCGA and METABRIC were used, one spans exons 11 and 12, and the other spans only exon 12. They most likely caught the pre-MIR21 gene and part of mature MIR21 as well (Appendix figure 6).

The use of different detection technology and detection of RNA products of VMP1 and MIR21 genes within cohorts may explain the discrepant

correlation values between VMP1 and hsa-miR-21-3p within cohort 1 and cohort 2 for TCGA and METABRIC/EGA.

Hsa-miR-21-3p correlated positively in TCGA with known drivers of this locus, like RPS6KB1 ($r = 0.67$, $p < 2.2 \times 10^{-16}$), PPM1D ($r = 0.58$, $p < 2.2 \times 10^{-16}$) and hsa-miR-21-5p ($r = 0.50$, $p = 1.75 \times 10^{-12}$). The correlation value among hsa-miR-21-3p with the above-mentioned genes in the METABRIC/EGA cohort was ($r = 0.11$, $p = 9.25 \times 10^{-05}$) with hsa-miR-21-5p, ($r = 0.37$, $p < 2.2 \times 10^{-16}$) with RPS6KB1, and ($r = 0.3$, $p < 2.2 \times 10^{-16}$) with PPM1D.

6.4 VMP1's function in HER2 positive cell lines

Here, I discuss data from an ongoing study, the purpose of which is to investigate the function of VMP1 in development of HER2 positive breast cancer cell lines. VMP1 is highly expressed in HER2 positive tumors, and due to its role in autophagy may play a role in resistance of HER2 positive tumors to trastuzumab.

VMP1 mRNA and protein levels were most highly expressed in ER⁺ and HER2⁺ cell lines, as with breast tumors. This was supported by analysis of VMP1 mRNA in the Broad Institute Breast cancer cell lines. The BT474 cell line expresses high levels of VMP1 and HER2. The assay set up involved silencing VMP1 on its own or simultaneously with HER2 and testing the effect in various traditional cell-based assays of gene oncogenicity. Silencing of VMP1 or ERBB2 genes did not affect each other's expression. The expression of the genes remained low for 96h after KD and at 120h was up to 50%. The BT474 cell line was established from a primary tumor in the breast gland and even though it expresses elevated levels of HER2, it grows slowly, with a doubling time of 96h.

Silencing of VMP1 did not associate with proliferation of HER2 positive cells. VMP1 mRNA also did not associate with Ki67 status of tumors within cohorts. Due to role of VMP1 in formation of zonula occludens, the effect of silencing VMP1 on the expression of ZO-1 and EMT1 was examined. Knocking down of VMP1 did not attenuate expression of E-CAD or ZO-1. Silencing of VMP1 did not affect induction of apoptosis. Perhaps immunofluorescence may be a better way to check the effect on tight junctions

Folkerts et al. found knocking down of VMP1, with a short hairpin RNA (shRNA) in primary leukemic cells and cell lines, had a strong effect on cell growth due to increased apoptosis and a decrease in survival and proliferation of cells (Folkerts et al., 2019). This is inconsistent with our

results in BT474 cell lines, where we detected HER2 expression at elevated levels; and silencing of VMP1 did not reduce proliferation due to the strong oncogenic effect of HER2.

Only two studies could be found that analyzed the functional role of VMP1 in breast cancer cell lines. In one study in MCF7 cells, VMP1 was knocked down with shRNA, which increased cell proliferation and migration (Sun et al., 2019). This result is opposite to what we would expect from our analyses in the patient cohorts where high VMP1 expression was associated with shorter survival. MCF7 cells express exceptionally low or no HER2, so the effect of VMP1, if any, would not be masked by HER2. The MCF7 cell line was established from a pleural effusion, so the cells have the characteristics of metastasis. They are rapidly growing cells and have multiple changes in their genome. Moreover, VMP1 may have different effects, depending on the environment. In our study, silencing VMP1 in BT474 cells did not have any effect on proliferation. The knockdown eliminated 92% of the protein, so inefficient knockdown is not the reason that no effect was observed. There could be several reasons why no effect was observed on proliferation. The BT474 cell line was derived from a primary tumor in the breast gland. Even if it expresses elevated levels of HER2, it grows slowly with a doubling time of 96 h. The cells were followed for a week after knockdown but VMP1 expression was observed 120h after KD and it is possible that it is not enough time to see the effect of VMP1. A more rapidly growing cell line may be necessary to observe any effect of VMP1. Alternatively, a different method, like shRNA, might produce a longer KD. We also realize that HER2 is a strong oncogene, which could possibly mask the effect of a weaker breast cancer gene like VMP1. It also is possible that VMP1 has no effect on proliferation. All the assays in which VMP1 was tested (such as for proliferation and apoptosis) were negative, i.e., VMP1 had no effect on its own or in conjunction with HER2.

Sun et al. also found silencing VMP1 decreased phosphorylation of ZO-1 and downregulated expression of E-Cad which was in total disagreement with our finding in BT474 cells (Sun et al., 2019). This discordance among our findings and theirs may be due to use of shRNA or using an HER2 negative cell line like MCF7 (their study). In HER2 positive cells, HER2 may confound the VMP1 effect due to HER2's strong oncogenic role. A clear example of this is the TOP2A gene, which is close to ERBB2. Moreover, topoisomerase II α is a target for anthracycline treatment (Engstrøm, Ytterhus, Vatten, Opdahl, & Bofin, 2014). Their co-amplification is a predictor of response to anthracycline treatment in breast cancer (Villman et al., 2006).

In another performed study in breast cancer cell lines by Sauermann et al. they found cell lines like MDA-MB-231 and HCC195 with high invasion capacity had significantly lower VMP1 mRNA quantity in comparison with cell lines like MCF7 and T-47D which had low invasion capacity with high VMP1 mRNA quantity (Sauermann et al., 2008). In our study MDA-MB-231 had lowest VMP1 mRNA in comparison with other breast cancer cell lines including T-47D and MCF7.

Experiments for investigation role of VMP1 in migration, invasion and resistance of HER2 positive cells to trastuzumab due to VMP1's function in autophagy were set up but could not perform due to time limitation.

6.5 Implications for future work

In this study we showed that screening of fusion genes in breast cancer cell lines and tumors can be used to identify genes with a role in tumor development.

BY quantifying levels of VMP1 mRNA in exploratory and validation cohorts and following results in two large data sets, we found VMP1 as a potential marker of poor prognosis in breast tumors and suggestive marker of poor prognosis within HER2 positive tumors. Including larger HER2 positive breast cancer cohorts with known treatment and response information, we might be able to shed light on a possible role for VMP1 in resistance of this aggressive subgroup to trastuzumab and lapatinib. Staining tumors with antibodies for VMP1 and autophagy proteins might unveil the connection between VMP1 upregulation and activation of autophagy in tumors.

In two study cohorts and a series of cell lines, we quantified hsa-miR-21-3p, which is processed from MIR21 that is transcribed from intron 10 of VMP1 but independently from it. The findings were followed up in larger breast cancer cohorts. The results suggest hsa-miR-21-3p is a marker of worse prognosis in breast cancer. Functional experiments focusing on hsa-miR-21-3p in breast cancer cell lines and with VMP1 are necessary to determine whether they are working together.

Due to interaction of VMP1 with the ZO-1 protein, it is most likely to affect migration and invasion, so studying its role on migration and metastasis is necessary.

Further investigations are necessary, to explore the role of VMP1 protein in tumor development, and in more breast cancer cell lines with respect to ER and HER2 receptor status. Due to VMP1's function in autophagy and its

known role in apoptosis of ovarian cancer (Zheng, Chen, Zhang, Zhan, & Chen, 2016), studying its role in trastuzumab resistance of HER2 positive breast cancer cells is important. Investigating the role of VMP1 in tumorigenesis and trastuzumab resistance in HER2 positive cell lines could not be performed due to time constraint. Important functional studies will test effects of HER2 receptor blockers and autophagy inhibitors on VMP1 expression within cell lines so as to discover improved combination therapy for this aggressive subtype of breast cancer.

Due to the slow growth rate of most HER2 positive cell lines, it would be necessary to change the gene silencing method and use short hairpin RNA instead of siRNA.

7 Conclusions

Taken together, the data presented suggested that analyzing genes involved in fusions may be a successful way to identify novel breast cancer genes.

High expression of VMP1 within breast tumors may be a marker of poor prognosis, particularly in tumors expressing HER2. Due to the role of VMP1 in the initiation of autophagy, high expression of VMP1 protein in HER2 positive tumors may activate autophagy and make tumors more prone to relapse.

High expression of hsa-miR-21-3p may be a marker of poor prognosis in breast tumors, irrespective of HER2, grade, and node status.

Given the circumstances, the data within this PhD thesis is the first report explaining how screening of fusion genes identified two genes that might associate with progression of breast cancer patients.

References

- Adams, B. D., Furneaux, H., & White, B. A. (2007). The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. *Mol Endocrinol*, *21*(5), 1132-1147. doi:10.1210/me.2007-0022
- Al-Bari, M. A. A. (2020). A current view of molecular dissection in autophagy machinery. *J Physiol Biochem*, *76*(3), 357-372. doi:10.1007/s13105-020-00746-0
- Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., . . . Sauter, G. (2004). Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res*, *64*(23), 8534-8540. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15574759
<http://cancerres.aacrjournals.org/content/canres/64/23/8534.full.pdf>
- Amirfallah, A., Arason, A., Einarsson, H., Gudmundsdottir, E. T., Freysteinsdottir, E. S., Olafsdottir, K. A., . . . Reynisdottir, I. (2019). High expression of the vacuole membrane protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer. *PLoS One*, *14*(8), e0221413. doi:10.1371/journal.pone.0221413
- Anandan, A., Sharifi, M., & O'Regan, R. (2020). Molecular Assays to Determine Optimal Duration of Adjuvant Endocrine Therapy in Breast Cancer. *Curr Treat Options Oncol*, *21*(10), 84. doi:10.1007/s11864-020-00788-y
- Anca Botezatu, I. V. I., Oana Popa, Adriana Plesa, Dana Manda, Irina Huica, Suzana Vladiu, Gabriela Anton and Corin Badiu. New Aspects in Molecular and Cellular Mechanisms of Human Carcinogenesis. In D. Bulgin (Ed.), *Mechanisms of Oncogene Activation* (2016 ed.): intechopen.
- Anca Botezatu, I. V. I., Oana Popa, Adriana Plesa, Dana Manda, Irina Huica, Suzana Vladiu, Gabriela Anton and Corin Badiu. (2016). Mechanisms of Oncogene Activation. In D. Bulgin (Ed.), *New Aspects in Molecular and Cellular Mechanisms of Human Carcinogenesis*: IntechOpen.

- Asmann, Y. W., Hossain, A., Necela, B. M., Middha, S., Kalari, K. R., Sun, Z., . . . Thompson, E. A. (2011). A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res*, 39(15), e100. doi:10.1093/nar/gkr362
- Asmann, Y. W., Necela, B. M., Kalari, K. R., Hossain, A., Baker, T. R., Carr, J. M., . . . Thompson, E. A. (2012). Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res*, 72(8), 1921-1928. doi:10.1158/0008-5472.can-11-3142
- Aure, M. R., Leivonen, S. K., Fleischer, T., Zhu, Q., Overgaard, J., Alsner, J., . . . Kristensen, V. N. (2013). Individual and combined effects of DNA methylation and copy number alterations on miRNA expression in breast tumors. *Genome Biol*, 14(11), R126. doi:10.1186/gb-2013-14-11-r126
- Babashah, S., & Soleimani, M. (2011). The oncogenic and tumour suppressive roles of microRNAs in cancer and apoptosis. *Eur J Cancer*, 47(8), 1127-1137. doi:10.1016/j.ejca.2011.02.008
- Babiceanu, M., Qin, F., Xie, Z., Jia, Y., Lopez, K., Janus, N., . . . Li, H. (2016). Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res*, 44(6), 2859-2872. doi:10.1093/nar/gkw032
- Baslan, T., Kendall, J., Volyanskyy, K., McNamara, K., Cox, H., D'Italia, S., . . . Hicks, J. (2020). Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *Elife*, 9. doi:10.7554/eLife.51480
- Bazzichetto, C., Conciatori, F., Pallocca, M., Falcone, I., Fanciulli, M., Cognetti, F., . . . Ciuffreda, L. (2019). PTEN as a Prognostic/Predictive Biomarker in Cancer: An Unfulfilled Promise? *Cancers (Basel)*, 11(4). doi:10.3390/cancers11040435
- Báez-Vega, P. M., Echevarría Vargas, I. M., Valiyeva, F., Encarnación-Rosado, J., Roman, A., Flores, J., . . . Vivas-Mejía, P. E. (2016). Targeting miR-21-3p inhibits proliferation and invasion of ovarian cancer cells. *Oncotarget*, 7(24), 36321-36337. doi:10.18632/oncotarget.9216
- Bertoli, G., Cava, C., & Castiglioni, I. (2015). MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer. *Theranostics*, 5(10), 1122-1143. doi:10.7150/thno.11543

- Bhat, A. A., Uppada, S., Achkar, I. W., Hashem, S., Yadav, S. K., Shanmugakonar, M., . . . Uddin, S. (2018). Tight Junction Proteins and Signaling Pathways in Cancer and Inflammation: A Functional Crosstalk. *Front Physiol*, *9*, 1942. doi:10.3389/fphys.2018.01942
- Bodily, W. R., Shirts, B. H., Walsh, T., Gulsuner, S., King, M. C., Parker, A., . . . Piccolo, S. R. (2020). Effects of germline and somatic events in candidate BRCA-like genes on breast-tumor signatures. *PLoS One*, *15*(9), e0239197. doi:10.1371/journal.pone.0239197
- Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis Part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit. *Br J Cancer*, *89*(4), 605-611. doi:10.1038/sj.bjc.6601120
- Britt, K. L., Cuzick, J., & Phillips, K. A. (2020). Key steps for effective breast cancer prevention. *Nat Rev Cancer*, *20*(8), 417-436. doi:10.1038/s41568-020-0266-x
- Broeks, A., Schmidt, M. K., Sherman, M. E., Couch, F. J., Hopper, J. L., Dite, G. S., . . . AOCS. (2011). Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*, *20*(16), 3289-3303. doi:10.1093/hmg/ddr228
- Calvo-Garrido, J., Carilla-Latorre, S., Lázaro-Diéguez, F., Egea, G., & Escalante, R. (2008). Vacuole membrane protein 1 is an endoplasmic reticulum protein required for organelle biogenesis, protein secretion, and development. *Mol Biol Cell*, *19*(8), 3442-3453. doi:10.1091/mbc.e08-01-0075
- Carbognin, L., Miglietta, F., Paris, I., & Dieci, M. V. (2019). Prognostic and Predictive Implications of PTEN in Breast Cancer: Unfulfilled Promises but Intriguing Perspectives. *Cancers (Basel)*, *11*(9). doi:10.3390/cancers11091401
- Cardoso, F., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rubio, I. T., . . . Senkus, E. (2019). Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol*, *30*(8), 1194-1220. doi:10.1093/annonc/mdz173
- Carlsson, E., Krohn, K., Ovaska, K., Lindberg, P., Häyry, V., Maliniemi, P., . . . Ranki, A. (2013). Neuron navigator 3 alterations in nervous system tumors associate with tumor malignancy grade and prognosis. *Genes Chromosomes Cancer*, *52*(2), 191-201. doi:10.1002/gcc.22019

- Carmona, F. J., Montemurro, F., Kannan, S., Rossi, V., Verma, C., Baselga, J., & Scaltriti, M. (2016). AKT signaling in ERBB2-amplified breast cancer. *Pharmacol Ther*, *158*, 63-70. doi:10.1016/j.pharmthera.2015.11.013
- Carpenter, R. L., & Lo, H. W. (2013). Regulation of Apoptosis by HER2 in Breast Cancer. *J Carcinog Mutagen*, *2013*(Suppl 7). doi:10.4172/2157-2518.S7-003
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, *2*(5), 401-404. doi:10.1158/2159-8290.cd-12-0095
- Céline Grandvallet, J. P. F., Franck Monnien, Gilles Despouy, Perez Valérie, Guittaut Michaël, Eric Hervouet and Paul Peixot. (2020). Autophagy is associated with a robust specific transcriptional signature in breast cancer subtypes. *Genes & Cancers*. doi:https://doi.org/10.18632/genesandcancer.208
- Chen, J., & Wang, X. (2014). MicroRNA-21 in breast cancer: diagnostic and prognostic potential. *Clin Transl Oncol*, *16*(3), 225-233. doi:10.1007/s12094-013-1132-z
- Chude, C. I., & Amaravadi, R. K. (2017). Targeting Autophagy in Cancer: Update on Clinical Trials and Novel Inhibitors. *Int J Mol Sci*, *18*(6). doi:10.3390/ijms18061279
- Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., . . . Network, T. R. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, *163*(2), 506-519. doi:10.1016/j.cell.2015.09.033
- Clare, S. E., & Shaw, P. L. (2016). "Big Data" for Breast Cancer: Where to look and what you will find. *NPJ Breast Cancer*, *2*, 16031-. doi:10.1038/npjbcancer.2016.31
- Colleoni, M., Sun, Z., Price, K. N., Karlsson, P., Forbes, J. F., Thürlimann, B., . . . Goldhirsch, A. (2016). Annual Hazard Rates of Recurrence for Breast Cancer During 24 Years of Follow-Up: Results From the International Breast Cancer Study Group Trials I to V. *J Clin Oncol*, *34*(9), 927-935. doi:10.1200/jco.2015.62.3504
- Comprehensive molecular portraits of human breast tumours. (2012). *Nature*, *490*(7418), 61-70. doi:10.1038/nature11412

- Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M., & Park, P. J. (2017). A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun*, *8*, 15180. doi:10.1038/ncomms15180
- Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., . . . Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), 346-352. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22522925<https://www.nature.com/articles/nature10983.pdf>
- Czech, B., & Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet*, *12*(1), 19-31. doi:10.1038/nrg2916
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*, *5*(10), 2929-2943.
- Doberstein, K., Bretz, N. P., Schirmer, U., Fiegl, H., Blaheta, R., Breunig, C., . . . Altevogt, P. (2014). miR-21-3p is a positive regulator of L1CAM in several human carcinomas. *Cancer Lett*, *354*(2), 455-466. doi:10.1016/j.canlet.2014.08.020
- Doherty, J., & Baehrecke, E. H. (2018). Life, death and autophagy. *Nat Cell Biol*, *20*(10), 1110-1117. doi:10.1038/s41556-018-0201-5
- Dower, C. M., Wills, C. A., Frisch, S. M., & Wang, H. G. (2018). Mechanisms and context underlying the role of autophagy in cancer metastasis. *Autophagy*, *14*(7), 1110-1128. doi:10.1080/15548627.2018.1450020
- Dudas, J., Ladanyi, A., Ingruber, J., Steinbichler, T. B., & Riechelmann, H. (2020). Epithelial to Mesenchymal Transition: A Mechanism that Fuels Cancer Radio/Chemoresistance. *Cells*, *9*(2). doi:10.3390/cells9020428
- Duffy, M. J., Synnott, N. C., & Crown, J. (2018). Mutant p53 in breast cancer: potential as a therapeutic target and biomarker. *Breast Cancer Res Treat*, *170*(2), 213-219. doi:10.1007/s10549-018-4753-7
- Duseti, N. J., Jiang, Y., Vaccaro, M. I., Tomasini, R., Azizi Samir, A., Calvo, E. L., . . . Iovanna, J. L. (2002). Cloning and expression of the rat vacuole membrane protein 1 (VMP1), a new gene activated in pancreas with acute pancreatitis, which promotes vacuole formation. *Biochem Biophys Res Commun*, *290*(2), 641-649. doi:10.1006/bbrc.2001.6244

- Easton, D. F., Pharoah, P. D., Antoniou, A. C., Tischkowitz, M., Tavtigian, S. V., Nathanson, K. L., . . . Foulkes, W. D. (2015). Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med*, *372*(23), 2243-2257. doi:10.1056/NEJMSr1501341
- Edgren, H., Murumagi, A., Kangaspeska, S., Nicorici, D., Hongisto, V., Kleivi, K., . . . Kallioniemi, O. (2011). Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol*, *12*(1), R6. doi:10.1186/gb-2011-12-1-r6
- Edwards, P. A., & Howarth, K. D. (2012). Are breast cancers driven by fusion genes? *Breast Cancer Res*, *14*(2), 303. doi:10.1186/bcr3122
- Eiriksdottir, G., Johannesdottir, G., Ingvarsson, S., Björnsdottir, I. B., Jonasson, J. G., Agnarsson, B. A., . . . Barkardottir, R. B. (1998). Mapping loss of heterozygosity at chromosome 13q: loss at 13q12-q13 is associated with breast tumour progression and poor prognosis. *Eur J Cancer*, *34*(13), 2076-2081. doi:10.1016/s0959-8049(98)00241-x
- Engstrøm, M. J., Ytterhus, B., Vatten, L. J., Opdahl, S., & Bofin, A. M. (2014). TOP2A gene copy number change in breast cancer. *J Clin Pathol*, *67*(5), 420-425. doi:10.1136/jclinpath-2013-202052
- Faltus, T., Yuan, J., Zimmer, B., Krämer, A., Loibl, S., Kaufmann, M., & Strebhardt, K. (2004). Silencing of the HER2/neu gene by siRNA inhibits proliferation and induces apoptosis in HER2/neu-overexpressing breast cancer cells. *Neoplasia*, *6*(6), 786-795. doi:10.1593/neo.04313
- Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A., & Fares, Y. (2020). Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduct Target Ther*, *5*(1), 28. doi:10.1038/s41392-020-0134-x
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., . . . Ren, G. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis*, *5*(2), 77-106. doi:10.1016/j.gendis.2018.05.001
- Feng, Y. H., & Tsao, C. J. (2016). Emerging role of microRNA-21 in cancer. *Biomed Rep*, *5*(4), 395-402. doi:10.3892/br.2016.747
- Fidler, I. J. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer*, *3*(6), 453-458. doi:10.1038/nrc1098

- Folkerts, H., Wierenga, A. T., van den Heuvel, F. A., Woldhuis, R. R., Kluit, D. S., Jaques, J., . . . Vellenga, E. (2019). Elevated VMP1 expression in acute myeloid leukemia amplifies autophagy and is protective against venetoclax-induced apoptosis. *Cell Death Dis*, *10*(6), 421. doi:10.1038/s41419-019-1648-4
- Frankel, L. B., Christoffersen, N. R., Jacobsen, A., Lindow, M., Krogh, A., & Lund, A. H. (2008). Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem*, *283*(2), 1026-1033. doi:10.1074/jbc.M707224200
- Fujita, S., Ito, T., Mizutani, T., Minoguchi, S., Yamamichi, N., Sakurai, K., & Iba, H. (2008). miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *J Mol Biol*, *378*(3), 492-504. doi:10.1016/j.jmb.2008.03.015
- Galluzzi, L., Pietrocola, F., Bravo-San Pedro, J. M., Amaravadi, R. K., Baehrecke, E. H., Cecconi, F., . . . Kroemer, G. (2015). Autophagy in malignant transformation and cancer progression. *EMBO J*, *34*(7), 856-880. doi:10.15252/embj.201490784
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, *6*(269), p11. doi:10.1126/scisignal.2004088
- Gao, Z., Liu, H., Shi, Y., Yin, L., Zhu, Y., & Liu, R. (2019). Identification of Cancer Stem Cell Molecular Markers and Effects of hsa-miR-21-3p on Stemness in Esophageal Squamous Cell Carcinoma. *Cancers (Basel)*, *11*(4). doi:10.3390/cancers11040518
- Gaur, A., Jewell, D. A., Liang, Y., Ridzon, D., Moore, J. H., Chen, C., . . . Israel, M. A. (2007). Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res*, *67*(6), 2456-2468. doi:10.1158/0008-5472.can-06-2698
- Giuliano, A. E., Edge, S. B., & Hortobagyi, G. N. (2018). Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Ann Surg Oncol*, *25*(7), 1783-1785. doi:10.1245/s10434-018-6486-6
- Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., . . . Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol*, *10*(5), 593-601. doi:10.1038/ncb1722

- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, *34*(Database issue), D140-144. doi:10.1093/nar/gkj112
- Gudmundsdottir, E. T., Barkardottir, R. B., Arason, A., Gunnarsson, H., Amundadottir, L. T., Agnarsson, B. A., . . . Reynisdottir, I. (2012). The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. *BMC Cancer*, *12*, 621. doi:10.1186/1471-2407-12-621
- Guo, L., Yang, L. Y., Fan, C., Chen, G. D., & Wu, F. (2012). Novel roles of Vmp1: inhibition metastasis and proliferation of hepatocellular carcinoma. *Cancer Sci*, *103*(12), 2110-2119. doi:10.1111/cas.12025
- Guo, Q., Wen, R., Shao, B., Li, Y., Jin, X., Deng, H., . . . Yu, F. (2018). Combined Let-7a and H19 Signature: A Prognostic Index of Progression-Free Survival in Primary Breast Cancer Patients. *J Breast Cancer*, *21*(2), 142-149. doi:10.4048/jbc.2018.21.2.142
- Guo, X. Z., Ye, X. L., Xiao, W. Z., Wei, X. N., You, Q. H., Che, X. H., . . . Yu, M. H. (2015). Downregulation of VMP1 confers aggressive properties to colorectal cancer. *Oncol Rep*, *34*(5), 2557-2566. doi:10.3892/or.2015.4240
- Hampton, O. A., Den Hollander, P., Miller, C. A., Delgado, D. A., Li, J., Coarfa, C., . . . Milosavljevic, A. (2009). A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res*, *19*(2), 167-177. doi:10.1101/gr.080259.108
- Hampton, O. A., Koriabine, M., Miller, C. A., Coarfa, C., Li, J., Den Hollander, P., . . . Milosavljevic, A. (2011). Long-range massively parallel mate pair sequencing detects distinct mutations and similar patterns of structural mutability in two breast cancer cell lines. *Cancer Genet*, *204*(8), 447-457. doi:10.1016/j.cancergen.2011.07.009
- Han, J., Hou, W., Lu, C., Goldstein, L. A., Stolz, D. B., Watkins, S. C., & Rabinowich, H. (2013). Interaction between Her2 and Beclin-1 proteins underlies a new mechanism of reciprocal regulation. *J Biol Chem*, *288*(28), 20315-20325. doi:10.1074/jbc.M113.461350
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

- Hanker, A. B., Sudhan, D. R., & Arteaga, C. L. (2020). Overcoming Endocrine Resistance in Breast Cancer. *Cancer Cell*, 37(4), 496-513. doi:10.1016/j.ccell.2020.03.009
- Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., . . . Cardoso, F. (2019). Breast cancer. *Nat Rev Dis Primers*, 5(1), 66. doi:10.1038/s41572-019-0111-2
- Hartsock, A., & Nelson, W. J. (2008). Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. *Biochim Biophys Acta*, 1778(3), 660-669. doi:10.1016/j.bbamem.2007.07.012
- Haverty, P. M., Fridlyand, J., Li, L., Getz, G., Beroukhim, R., Lohr, S., . . . Chant, J. (2008). High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer*, 47(6), 530-542. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18335499
<https://onlinelibrary.wiley.com/doi/pdf/10.1002/gcc.20558>
- Hemmatzadeh, M., Mohammadi, H., Jadidi-Niaragh, F., Asghari, F., & Yousefi, M. (2016). The role of oncomirs in the pathogenesis and treatment of breast cancer. *Biomed Pharmacother*, 78, 129-139. doi:10.1016/j.biopha.2016.01.026
- Hoebeek, J., van der Luijt, R., Poppe, B., De Smet, E., Yigit, N., Claes, K., . . . Vandesompele, J. (2005). Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab Invest*, 85(1), 24-33. doi:10.1038/labinvest.3700209
- Holleczek, B., Stegmaier, C., Radosa, J. C., Solomayer, E. F., & Brenner, H. (2019). Risk of loco-regional recurrence and distant metastases of patients with invasive breast cancer up to ten years after diagnosis - results from a registry-based study from Germany. *BMC Cancer*, 19(1), 520. doi:10.1186/s12885-019-5710-5
- Holz, M. K. (2012). The role of S6K1 in ER-positive breast cancer. *Cell Cycle*, 11(17), 3159-3165. doi:10.4161/cc.21194
- Hossain, A., Kuo, M. T., & Saunders, G. F. (2006). Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. *Mol Cell Biol*, 26(21), 8191-8201. doi:10.1128/mcb.00242-06

- Hou, N., Guo, Z., Zhao, G., Jia, G., Luo, B., Shen, X., & Bai, Y. (2018). Inhibition of microRNA-21-3p suppresses proliferation as well as invasion and induces apoptosis by targeting RNA-binding protein with multiple splicing through Smad4/extra cellular signal-regulated protein kinase signalling pathway in human colorectal cancer HCT116 cells. *Clin Exp Pharmacol Physiol*, *45*(7), 729-741. doi:10.1111/1440-1681.12931
- Howlader, N., Altekruse, S. F., Li, C. I., Chen, V. W., Clarke, C. A., Ries, L. A., & Cronin, K. A. (2014). US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst*, *106*(5). doi:10.1093/jnci/dju055
- Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., . . . Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, *7*, 96. doi:10.1186/1471-2164-7-96
- Hudis, C. A. (2007). Trastuzumab--mechanism of action and use in clinical practice. *N Engl J Med*, *357*(1), 39-51. doi:10.1056/NEJMra043186
- The Human Protein Atlas (<https://www.proteinatlas.org/>). Retrieved from <https://www.proteinatlas.org>
- Imani, S., Wu, R. C., & Fu, J. (2018). MicroRNA-34 family in breast cancer: from research to therapeutic potential. *J Cancer*, *9*(20), 3765-3775. doi:10.7150/jca.25576
- Inaki, K., Hillmer, A. M., Ukil, L., Yao, F., Woo, X. Y., Vardy, L. A., . . . Liu, E. T. (2011). Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome Res*, *21*(5), 676-687. doi:10.1101/gr.113225.110
- Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., . . . Croce, C. M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*, *65*(16), 7065-7070. doi:10.1158/0008-5472.can-05-1783
- Janser, F. A., Tschan, M. P., & Langer, R. (2019). The role of autophagy in HER2-targeted therapy. *Swiss Med Wkly*, *149*, w20138. doi:10.4414/smw.2019.20138
- Januškevičienė, I., & Petrikaitė, V. (2019). Heterogeneity of breast cancer: The importance of interaction between different tumor cell populations. *Life Sci*, *239*, 117009. doi:10.1016/j.lfs.2019.117009

- Jia, W., Qiu, K., He, M., Song, P., Zhou, Q., Zhou, F., . . . Guo, G. (2013). SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol*, *14*(2), R12. doi:10.1186/gb-2013-14-2-r12
- Jiang, H., & Rugo, H. S. (2015). Human epidermal growth factor receptor 2 positive (HER2+) metastatic breast cancer: how the latest results are improving therapeutic options. *Ther Adv Med Oncol*, *7*(6), 321-339. doi:10.1177/1758834015599389
- Jiang, M., Zhang, P., Hu, G., Xiao, Z., Xu, F., Zhong, T., . . . Zhang, W. (2013). Relative expressions of miR-205-5p, miR-205-3p, and miR-21 in tissues and serum of non-small cell lung cancer patients. *Mol Cell Biochem*, *383*(1-2), 67-75. doi:10.1007/s11010-013-1755-y
- Jiao, W., Leng, X., Zhou, Q., Wu, Y., Sun, L., Tan, Y., . . . Li, J. (2017). Different miR-21-3p isoforms and their different features in colorectal cancer. *Int J Cancer*, *141*(10), 2103-2111. doi:10.1002/ijc.30902
- Jolly, C., & Van Loo, P. (2018). Timing somatic events in the evolution of cancer. *Genome Biol*, *19*(1), 95. doi:10.1186/s13059-018-1476-3
- Jonsson, G., Staaf, J., Vallon-Christersson, J., Ringner, M., Gruvberger-Saal, S. K., Saal, L. H., . . . Borg, A. (2012). The retinoblastoma gene undergoes rearrangements in BRCA1-deficient basal-like breast cancer. *Cancer Res*, *72*(16), 4028-4036. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22706203<http://cancerres.aacrjournals.org/content/canres/72/16/4028.full.pdf>
- Jonsson, G., Staaf, J., Vallon-Christersson, J., Ringner, M., Holm, K., Hegardt, C., . . . Borg, A. (2010). Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res*, *12*(3), R42. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20576095<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917037/pdf/bcr2596.pdf>
- Joy, A. A., & Mackey, J. R. (2006). Adjuvant trastuzumab: progress, controversies, and the steps ahead. *Curr Oncol*, *13*(1), 8-13.

- Kalyana-Sundaram, S., Shankar, S., Deroo, S., Iyer, M. K., Palanisamy, N., Chinnaiyan, A. M., & Kumar-Sinha, C. (2012). Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia*, *14*(8), 702-708. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22952423>https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431177/pdf/neo1408_0702.pdf
- Kampf, C., Olsson, I., Ryberg, U., Sjöstedt, E., & Pontén, F. (2012). Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *J Vis Exp*(63). doi:10.3791/3620
- Kangaspeska, S., Hultsch, S., Edgren, H., Nicorici, D., Murumägi, A., & Kallioniemi, O. (2012). Reanalysis of RNA-sequencing data reveals several additional fusion genes with multiple isoforms. *PLoS One*, *7*(10), e48745. doi:10.1371/journal.pone.0048745
- Karantza-Wadsworth, V., Patel, S., Kravchuk, O., Chen, G., Mathew, R., Jin, S., & White, E. (2007). Autophagy mitigates metabolic stress and genome damage in mammary tumorigenesis. *Genes Dev*, *21*(13), 1621-1635. doi:10.1101/gad.1565707
- Karenko, L., Hahtola, S., Päivinen, S., Karhu, R., Syrjä, S., Kähkönen, M., . . . Ranki, A. (2005). Primary cutaneous T-cell lymphomas show a deletion or translocation affecting NAV3, the human UNC-53 homologue. *Cancer Res*, *65*(18), 8101-8110. doi:10.1158/0008-5472.Can-04-0366
- Kato, M., Paranjape, T., Müller, R. U., Nallur, S., Gillespie, E., Keane, K., . . . Slack, F. J. (2009). The mir-34 microRNA is required for the DNA damage response in vivo in *C. elegans* and in vitro in human breast cancer cells. *Oncogene*, *28*(25), 2419-2424. doi:10.1038/onc.2009.106
- Kim, J., Kim, S., Ko, S., In, Y. H., Moon, H. G., Ahn, S. K., . . . Han, W. (2015). Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. *Genes Chromosomes Cancer*, *54*(11), 681-691. doi:10.1002/gcc.22279
- King, C. R., Kraus, M. H., & Aaronson, S. A. (1985). Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science*, *229*(4717), 974-976. doi:10.1126/science.2992089
- Klijn, C., Durinck, S., Stawiski, E. W., Haverty, P. M., Jiang, Z., Liu, H., . . . Zhang, Z. (2015). A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol*, *33*(3), 306-312. doi:10.1038/nbt.3080

- Klinge, C. M. (2018). Non-Coding RNAs in Breast Cancer: Intracellular and Intercellular Communication. *Noncoding RNA*, 4(4). doi:10.3390/ncrna4040040
- Krabbameinsskrá. Retrieved from <https://www.krabb.is/krabbameinsskra>
- Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., . . . Olsson, H. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*, 317(23), 2402-2416. doi:10.1001/jama.2017.7112
- Kufe DW, P. R., Weichselbaum RR, et al. (2003). *Holland-Frei Cancer Medicine. 6th edition*
Hamilton (ON): BC Decker.
- Kumar-Sinha, C., Kalyana-Sundaram, S., & Chinnaiyan, A. M. (2015). Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med*, 7, 129. doi:10.1186/s13073-015-0252-1
- Lakhani SR, E. I., Schnitt SJ, Tan PH, van de Vijver MJ. (2012). *WHO Classification of Tumours of the Breast* (Vol. 4).
- Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*, 15(3), 178-196. doi:10.1038/nrm3758
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5), 843-854. doi:10.1016/0092-8674(93)90529-y
- Lee, Y. M., Oh, M. H., Go, J. H., Han, K., & Choi, S. Y. (2020). Molecular subtypes of triple-negative breast cancer: understanding of subtype categories and clinical implication. *Genes Genomics*, 42(12), 1381-1387. doi:10.1007/s13258-020-01014-7
- Levy, J. M. M., Towers, C. G., & Thorburn, A. (2017). Targeting autophagy in cancer. *Nat Rev Cancer*, 17(9), 528-542. doi:10.1038/nrc.2017.53
- Li, X., He, S., & Ma, B. (2020). Autophagy and autophagy-related proteins in cancer. *Mol Cancer*, 19(1), 12. doi:10.1186/s12943-020-1138-4
- Li, Y., Yang, D., Yin, X., Zhang, X., Huang, J., Wu, Y., . . . Ren, G. (2020). Clinicopathological Characteristics and Breast Cancer-Specific Survival of Patients With Single Hormone Receptor-Positive Breast Cancer. *JAMA Netw Open*, 3(1), e1918160. doi:10.1001/jamanetworkopen.2019.18160
- Litovchick, L. (2018). Preparing Whole-Cell Lysates for Immunoblotting. *Cold Spring Harb Protoc*, 2018(7). doi:10.1101/pdb.prot098400

- Little, P., Lin, D. Y., & Sun, W. (2019). Associating somatic mutations to clinical outcomes: a pan-cancer study of survival time. *Genome Med*, 11(1), 37. doi:10.1186/s13073-019-0643-9
- Liu, T., Zhao, L., Chen, W., Li, Z., Hou, H., Ding, L., & Li, X. (2014). Inactivation of von Hippel-Lindau increases ovarian cancer cell aggressiveness through the HIF1 α /miR-210/VMP1 signaling pathway. *Int J Mol Med*, 33(5), 1236-1242. doi:10.3892/ijmm.2014.1661
- Liu, Y., Xu, J., Choi, H. H., Han, C., Fang, Y., Li, Y., . . . Zhang, X. (2018). Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer. *Nat Commun*, 9(1), 4718. doi:10.1038/s41467-018-07264-0
- Lo Ré, A. E., Fernández-Barrena, M. G., Almada, L. L., Mills, L. D., Elswa, S. F., Lund, G., . . . Fernandez-Zapico, M. E. (2012). Novel AKT1-GLI3-VMP1 pathway mediates KRAS oncogene-induced autophagy in cancer cells. *J Biol Chem*, 287(30), 25325-25334. doi:10.1074/jbc.M112.370809
- Lo, T. F., Tsai, W. C., & Chen, S. T. (2013). MicroRNA-21-3p, a berberine-induced miRNA, directly down-regulates human methionine adenosyltransferases 2A and 2B and inhibits hepatoma cell growth. *PloS one*, 8(9), e75628. doi:10.1371/journal.pone.0075628
- Low, S. K., Zembutsu, H., & Nakamura, Y. (2018). Breast cancer: The translation of big genomic data to cancer precision medicine. *Cancer Sci*, 109(3), 497-506. doi:10.1111/cas.13463
- Lu, Z. M., Lin, Y. F., Jiang, L., Chen, L. S., Luo, X. N., Song, X. H., . . . Zhang, S. Y. (2014). Micro-ribonucleic acid expression profiling and bioinformatic target gene analyses in laryngeal carcinoma. *Oncotargets Ther*, 7, 525-533. doi:10.2147/ott.S59871
- Luo, L. J., Yang, F., Ding, J. J., Yan, D. L., Wang, D. D., Yang, S. J., . . . Tang, J. H. (2016). MiR-31 inhibits migration and invasion by targeting SATB2 in triple negative breast cancer. *Gene*, 594(1), 47-58. doi:10.1016/j.gene.2016.08.057
- Ma, L., Teruya-Feldstein, J., & Weinberg, R. A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, 449(7163), 682-688. doi:10.1038/nature06174
- Maes, H., Rubio, N., Garg, A. D., & Agostinis, P. (2013). Autophagy: shaping the tumor microenvironment and therapeutic response. *Trends Mol Med*, 19(7), 428-446. doi:10.1016/j.molmed.2013.04.005

- Malhotra, G. K., Zhao, X., Band, H., & Band, V. (2010). Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther*, 10(10), 955-960. doi:10.4161/cbt.10.10.13879
- Masoud, V., & Pagès, G. (2017). Targeted therapies in breast cancer: New challenges to fight against resistance. *World J Clin Oncol*, 8(2), 120-134. doi:10.5306/wjco.v8.i2.120
- Mathew, R., Kongara, S., Beaudoin, B., Karp, C. M., Bray, K., Degenhardt, K., . . . White, E. (2007). Autophagy suppresses tumor progression by limiting chromosomal instability. *Genes Dev*, 21(11), 1367-1381. doi:10.1101/gad.1545107
- Mele, L., Del Vecchio, V., Liccardo, D., Prisco, C., Schwerdtfeger, M., Robinson, N., . . . La Noce, M. (2020). The role of autophagy in resistance to targeted therapies. *Cancer Treat Rev*, 88, 102043. doi:10.1016/j.ctrv.2020.102043
- Mertens, F., Johansson, B., Fioretos, T., & Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nat Rev Cancer*, 15(6), 371-381. doi:10.1038/nrc3947
- Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., . . . Easton, D. F. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet*, 47(4), 373-380. doi:10.1038/ng.3242
- Mirabelli, P., Coppola, L., & Salvatore, M. (2019). Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancers (Basel)*, 11(8). doi:10.3390/cancers11081098
- Molejon, M. I., Ropolo, A., Re, A. L., Boggio, V., & Vaccaro, M. I. (2013). The VMP1-Becn1 interaction regulates autophagy induction. *Sci Rep*, 3, 1055. doi:10.1038/srep01055
- Molejon, M. I., Ropolo, A., & Vaccaro, M. I. (2013). VMP1 is a new player in the regulation of the autophagy-specific phosphatidylinositol 3-kinase complex activation. *Autophagy*, 9(6), 933-935. doi:10.4161/auto.24390
- Mollon, L. E., Anderson, E. J., Dean, J. L., Warholak, T. L., Aizer, A., Platt, E. A., . . . Davis, L. E. (2020). A Systematic Literature Review of the Prognostic and Predictive Value of PIK3CA Mutations in HR(+)/HER2(-) Metastatic Breast Cancer. *Clin Breast Cancer*, 20(3), e232-e243. doi:10.1016/j.clbc.2019.08.011

- Morishita, H., Zhao, Y. G., Tamura, N., Nishimura, T., Kanda, Y., Sakamaki, Y., . . . Mizushima, N. (2019). A critical role of VMP1 in lipoprotein secretion. *Elife*, *8*. doi:10.7554/eLife.48834
- Mowers, E. E., Sharifi, M. N., & Macleod, K. F. (2017). Autophagy in cancer metastasis. *Oncogene*, *36*(12), 1619-1630. doi:10.1038/onc.2016.333
- N Hamajima, T. R., CM Friedenreich, SM Gapstur, MM Gaudet, RJ Coates, JM Liff, E Kakouri, Y Marcou. (2019). Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet*, *394*(10204), 1159-1168. doi:10.1016/s0140-6736(19)31709-x
- Natrajan, R., Lambros, M. B., Rodríguez-Pinilla, S. M., Moreno-Bueno, G., Tan, D. S., Marchió, C., . . . Reis-Filho, J. S. (2009). Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res*, *15*(8), 2711-2722. doi:10.1158/1078-0432.Ccr-08-1878
- Network, C. G. A. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61-70.
- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., . . . Gray, J. W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, *10*(6), 515-527.
Retrieved from
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17157791
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2730521/pdf/nihms122151.pdf>
- Nik-Zainal. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, *534*(7605), 47-54.
doi:10.1038/nature17676
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., . . . Stratton, M. R. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, *534*(7605), 47-54.
doi:10.1038/nature17676
- Noh, W. C., Kim, Y. H., Kim, M. S., Koh, J. S., Kim, H. A., Moon, N. M., & Paik, N. S. (2008). Activation of the mTOR signaling pathway in breast cancer and its correlation with the clinicopathologic variables. *Breast Cancer Res Treat*, *110*(3), 477-483. doi:10.1007/s10549-007-9746-x
- Nwabo Kamdje, A. H., Seke Etet, P. F., Vecchio, L., Muller, J. M., Krampera, M., & Lukong, K. E. (2014). Signaling pathways in breast cancer: therapeutic targeting of the microenvironment. *Cell Signal*, *26*(12), 2843-2856. doi:10.1016/j.cellsig.2014.07.034

- O'Bryan, S., Dong, S., Mathis, J. M., & Alahari, S. K. (2017). The roles of oncogenic miRNAs and their therapeutic importance in breast cancer. *Eur J Cancer*, *72*, 1-11. doi:10.1016/j.ejca.2016.11.004
- O'Day, E., & Lal, A. (2010). MicroRNAs and their target gene networks in breast cancer. *Breast Cancer Res*, *12*(2), 201. doi:10.1186/bcr2484
- Ohshima, K., Hatakeyama, K., Nagashima, T., Watanabe, Y., Kanto, K., Doi, Y., . . . Yamaguchi, K. (2017). Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Sci Rep*, *7*(1), 641. doi:10.1038/s41598-017-00219-3
- Ouyang, M., Li, Y., Ye, S., Ma, J., Lu, L., Lv, W., . . . Wang, W. (2014). MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. *PloS one*, *9*(5), e96228. doi:10.1371/journal.pone.0096228
- Pan, H., Gray, R., Braybrooke, J., Davies, C., Taylor, C., McGale, P., . . . Hayes, D. F. (2017). 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N Engl J Med*, *377*(19), 1836-1846. doi:10.1056/NEJMoa1701830
- Paratala, B. S., Dolfi, S. C., Khiabani, H., Rodriguez-Rodriguez, L., Ganesan, S., & Hirshfield, K. M. (2016). Emerging Role of Genomic Rearrangements in Breast Cancer: Applying Knowledge from Other Cancers. *Biomark Cancer*, *8*(Supple 1), 1-14. doi:10.4137/bic.s34417
- Paul, D. (2020). The systemic hallmarks of cancer *Cancer Metastasis Treatment*, *6*(29). doi:10.20517/2394-4722.2020.63
- Pereira, B., Chin, S. F., Rueda, O. M., Vollan, H. K., Provenzano, E., Bardwell, H. A., . . . Caldas, C. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*, *7*, 11479. doi:10.1038/ncomms11479
- Persson, H., Søkilde, R., Häkkinen, J., Pirona, A. C., Vallon-Christersson, J., Kvist, A., . . . Rovira, C. (2017). Frequent miRNA-convergent fusion gene events in breast cancer. *Nat Commun*, *8*(1), 788. doi:10.1038/s41467-017-01176-1
- Persson, M., Andrén, Y., Mark, J., Horlings, H. M., Persson, F., & Stenman, G. (2009). Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc Natl Acad Sci U S A*, *106*(44), 18740-18744. doi:10.1073/pnas.0909114106

- Pérez-Tenorio, G., Karlsson, E., Waltersson, M. A., Olsson, B., Holmlund, B., Nordenskjöld, B., . . . Stål, O. (2011). Clinical potential of the mTOR targets S6K1 and S6K2 in breast cancer. *Breast Cancer Res Treat*, 128(3), 713-723. doi:10.1007/s10549-010-1058-x
- Pfeffer, C. M., & Singh, A. T. K. (2018). Apoptosis: A Target for Anticancer Therapy. *Int J Mol Sci*, 19(2). doi:10.3390/ijms19020448
- Pink, R. C., Samuel, P., Massa, D., Caley, D. P., Brooks, S. A., & Carter, D. R. (2015). The passenger strand, miR-21-3p, plays a role in mediating cisplatin resistance in ovarian cancer cells. *Gynecol Oncol*, 137(1), 143-151. doi:10.1016/j.ygyno.2014.12.042
- Pohlmann, P. R., Mayer, I. A., & Mernaugh, R. (2009). Resistance to Trastuzumab in Breast Cancer. *Clin Cancer Res*, 15(24), 7479-7491. doi:10.1158/1078-0432.Ccr-09-0636
- Provenzano, E., Ulaner, G. A., & Chin, S. F. (2018). Molecular Classification of Breast Cancer. *PET Clin*, 13(3), 325-338. doi:10.1016/j.cpet.2018.02.004
- Qian, Q., Zhou, H., Chen, Y., Shen, C., He, S., Zhao, H., . . . Gu, W. (2014). VMP1 related autophagy and apoptosis in colorectal cancer cells: VMP1 regulates cell death. *Biochem Biophys Res Commun*, 443(3), 1041-1047. doi:10.1016/j.bbrc.2013.12.090
- Qu, A., Du, L., Yang, Y., Liu, H., Li, J., Wang, L., . . . Wang, C. (2014). Hypoxia-inducible MiR-210 is an independent prognostic factor and contributes to metastasis in colorectal cancer. *PLoS One*, 9(3), e90952. doi:10.1371/journal.pone.0090952
- Quintás-Cardama, A., & Cortes, J. (2009). Molecular biology of bcr-abl1-positive chronic myeloid leukemia. *Blood*, 113(8), 1619-1630. doi:10.1182/blood-2008-03-144790
- Quintás-Cardama, A., Kantarjian, H., & Cortes, J. (2009). Imatinib and beyond--exploring the full potential of targeted therapy for CML. *Nat Rev Clin Oncol*, 6(9), 535-543. doi:10.1038/nrclinonc.2009.112
- Rakha, E. A., Reis-Filho, J. S., Baehner, F., Dabbs, D. J., Decker, T., Eusebi, V., . . . Ellis, I. O. (2010). Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*, 12(4), 207. doi:10.1186/bcr2607
- Rakha, E. A., Reis-Filho, J. S., & Ellis, I. O. (2010). Combinatorial biomarker expression in breast cancer. *Breast Cancer Res Treat*, 120(2), 293-308. doi:10.1007/s10549-010-0746-x

- Ramroop, J. R., Gerber, M. M., & Toland, A. E. (2019). Germline Variants Impact Somatic Events during Tumorigenesis. *Trends Genet*, 35(7), 515-526. doi:10.1016/j.tig.2019.04.005
- Ranki, A., Väkevä, L., Sipilä, L., & Krohn, K. (2011). Molecular markers associated with clinical response to bexarotene therapy in cutaneous T-cell lymphoma. *Acta Derm Venereol*, 91(5), 568-573. doi:10.2340/00015555-1114
- Renoir, J. M., Marsaud, V., & Lazennec, G. (2013). Estrogen receptor signaling as a target for novel breast cancer therapeutics. *Biochem Pharmacol*, 85(4), 449-465. doi:10.1016/j.bcp.2012.10.018
- Reynisdottir, I., Arason, A., Einarsdottir, B. O., Gunnarsson, H., Staaf, J., Vallon-Christersson, J., . . . Barkardottir, R. B. (2013). High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. *Cancer Med*, 2(4), 437-446. doi:10.1002/cam4.88
- Ribas, J., & Lupold, S. E. (2010). The transcriptional regulation of miR-21, its multiple transcripts, and their implication in prostate cancer. *Cell Cycle*, 9(5), 923-929. doi:10.4161/cc.9.5.10930
- Ribas, J., Ni, X., Castanares, M., Liu, M. M., Esopi, D., Yegnasubramanian, S., . . . Lupold, S. E. (2012). A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res*, 40(14), 6821-6833. doi:10.1093/nar/gks308
- Riggs, N., Aguet, M., & Stamenkovic, I. (2018). Cancer Metastasis: A Reappraisal of Its Underlying Mechanisms and Their Relevance to Treatment. *Annu Rev Pathol*, 13, 117-140. doi:10.1146/annurev-pathol-020117-044127
- Robinson, D. R., Kalyana-Sundaram, S., Wu, Y. M., Shankar, S., Cao, X., Ateeq, B., . . . Chinnaiyan, A. M. (2011). Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med*, 17(12), 1646-1651. doi:10.1038/nm.2580
- Roskoski, R., Jr. (2014). The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol Res*, 79, 34-74. doi:10.1016/j.phrs.2013.11.002
- Rueda, O. M., Sammut, S. J., Seoane, J. A., Chin, S. F., Caswell-Jin, J. L., Callari, M., . . . Curtis, C. (2019). Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*, 567(7748), 399-404. doi:10.1038/s41586-019-1007-8

- Russo, M., & Russo, G. L. (2018). Autophagy inducers in cancer. *Biochem Pharmacol*, 153, 51-61. doi:10.1016/j.bcp.2018.02.007
- Salvador, E., Burek, M., & Förster, C. Y. (2016). Tight Junctions and the Tumor Microenvironment. *Curr Pathobiol Rep*, 4, 135-145. doi:10.1007/s40139-016-0106-6
- Santana-Codina, N., Mancias, J. D., & Kimmelman, A. C. (2017). The Role of Autophagy in Cancer. *Annu Rev Cancer Biol*, 1, 19-39. doi:10.1146/annurev-cancerbio-041816-122338
- Sauermann, M., Sahin, O., Sülthmann, H., Hahne, F., Blaszkiewicz, S., Majety, M., . . . Arlt, D. (2008). Reduced expression of vacuole membrane protein 1 affects the invasion capacity of tumor cells. *Oncogene*, 27(9), 1320-1326. doi:10.1038/sj.onc.1210743
- Schulte, I., Batty, E. M., Pole, J. C., Blood, K. A., Mo, S., Cooke, S. L., . . . Edwards, P. A. (2012). Structural analysis of the genome of breast cancer cell line ZR-75-30 identifies twelve expressed fusion genes. *BMC Genomics*, 13, 719. doi:10.1186/1471-2164-13-719
- Selcuklu, S. D., Donoghue, M. T., & Spillane, C. (2009). miR-21 as a key regulator of oncogenic processes. *Biochem Soc Trans*, 37(Pt 4), 918-925. doi:10.1042/bst0370918
- Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., . . . Fan, X. (2019). Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet*, 20(1), 175. doi:10.1186/s12881-019-0909-5
- Shiu, K. K., Natrajan, R., Geyer, F. C., Ashworth, A., & Reis-Filho, J. S. (2010). DNA amplifications in breast cancer: genotypic-phenotypic correlations. *Future Oncol*, 6(6), 967-984. doi:10.2217/fon.10.56
- Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J Clin*, 70(1), 7-30. doi:10.3322/caac.21590
- Silva, G. O., He, X., Parker, J. S., Gatz, M. L., Carey, L. A., Hou, J. P., . . . Perou, C. M. (2015). Cross-species DNA copy number analyses identifies multiple 1q21-q23 subtype-specific driver genes for breast cancer. *Breast Cancer Res Treat*, 152(2), 347-356. doi:10.1007/s10549-015-3476-2
- Singh, S. S., Vats, S., Chia, A. Y., Tan, T. Z., Deng, S., Ong, M. S., . . . Kumar, A. P. (2018). Dual role of autophagy in hallmarks of cancer. *Oncogene*, 37(9), 1142-1158. doi:10.1038/s41388-017-0046-6

- Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., . . . Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*, *344*(11), 783-792.
doi:10.1056/nejm200103153441101
- Staaf, J., Jonsson, G., Ringner, M., Vallon-Christersson, J., Grabau, D., Arason, A., . . . Borg, A. (2010). High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res*, *12*(3), R25. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20459607<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917012/pdf/bcr2568.pdf>
- Stavast, C. J., & Erkeland, S. J. (2019). The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. *Cells*, *8*(11).
doi:10.3390/cells8111465
- Stephens, P. J., McBride, D. J., Lin, M. L., Varela, I., Pleasance, E. D., Simpson, J. T., . . . Stratton, M. R. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*, *462*(7276), 1005-1010. doi:10.1038/nature08645
- Sun, J. W., Yang, J. Y., Chen, X., Li, L., Hao, J., Niu, H., . . . Qiu, S. (2019). Silencing of VMP1 makes breast cancer cells more aggressive and resistant to 5-Fu. *International Journal of Clinical and Experimental Medicine*, *12*(11), 13096-13107. Retrieved from <Go to ISI>://WOS:000500300700062
- Suvanto, M., Beesley, J., Blomqvist, C., Chenevix-Trench, G., Khan, S., & Nevanlinna, H. (2020). SNPs in lncRNA Regions and Breast Cancer Risk. *Front Genet*, *11*, 550. doi:10.3389/fgene.2020.00550
- Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., . . . Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, *98*(19), 10869-10874. doi:10.1073/pnas.191367098
- Tavazoie, S. F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P. D., . . . Massagué, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*, *451*(7175), 147-152.
doi:10.1038/nature06487

- Tábara, L. C., Vicente, J. J., Biazik, J., Eskelinen, E. L., Vincent, O., & Escalante, R. (2018). Vacuole membrane protein 1 marks endoplasmic reticulum subdomains enriched in phospholipid synthesizing enzymes and is required for phosphoinositide distribution. *Traffic*, *19*(8), 624-638. doi:10.1111/tra.12581
- Tian, X., Gu, T., Patel, S., Bode, A. M., Lee, M. H., & Dong, Z. (2019). CRISPR/Cas9 - An evolving biological tool kit for cancer biology and oncology. *NPJ Precis Oncol*, *3*, 8. doi:10.1038/s41698-019-0080-7
- Tognon, C., Knezevich, S. R., Huntsman, D., Roskelley, C. D., Melnyk, N., Mathers, J. A., . . . Sorensen, P. H. (2002). Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell*, *2*(5), 367-376. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12450792>
- Tower, H., Ruppert, M., & Britt, K. (2019). The Immune Microenvironment of Breast Cancer Progression. *Cancers (Basel)*, *11*(9). doi:10.3390/cancers11091375
- Tung, N. M., & Garber, J. E. (2018). BRCA1/2 testing: therapeutic implications for breast cancer management. *Br J Cancer*, *119*(2), 141-152. doi:10.1038/s41416-018-0127-5
- U.S. Food and Drug Administration , O. C. o. E., Center for Drug Evaluation and Research (CDER), and Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration. (December 2018). *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry* United State of America: U.S. Department of Health and Human Services Food and Drug Administration Oncology Center of Excellence Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)
- Vaccaro, M. I., Ropolo, A., Grasso, D., & Iovanna, J. L. (2008). A novel mammalian trans-membrane protein reveals an alternative initiation pathway for autophagy. *Autophagy*, *4*(3), 388-390. doi:10.4161/auto.5656
- Valastyan, S., Reinhardt, F., Benaich, N., Calogrias, D., Szász, A. M., Wang, Z. C., . . . Weinberg, R. A. (2009). A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell*, *137*(6), 1032-1046. doi:10.1016/j.cell.2009.03.047

- van der Hage, J. A., van den Broek, L. J., Legrand, C., Clahsen, P. C., Bosch, C. J., Robanus-Maandag, E. C., . . . van de Vijver, M. J. (2004). Overexpression of P70 S6 kinase protein is associated with increased risk of locoregional recurrence in node-negative premenopausal early breast cancer patients. *Br J Cancer*, *90*(8), 1543-1550. doi:10.1038/sj.bjc.6601741
- Varley, K. E., Gertz, J., Roberts, B. S., Davis, N. S., Bowling, K. M., Kirby, M. K., . . . Myers, R. M. (2014). Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat*, *146*(2), 287-297. doi:10.1007/s10549-014-3019-2
- Veeraraghavan, J., Ma, J., Hu, Y., & Wang, X. S. (2016). Recurrent and pathological gene fusions in breast cancer: current advances in genomic discovery and clinical implications. *Breast Cancer Res Treat*, *158*(2), 219-232. doi:10.1007/s10549-016-3876-y
- Veeraraghavan, J., Tan, Y., Cao, X. X., Kim, J. A., Wang, X., Chamness, G. C., . . . Wang, X. S. (2014). Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun*, *5*, 4577. doi:10.1038/ncomms5577
- Vega-Rubín-de-Celis, S. (2019). The Role of Beclin 1-Dependent Autophagy in Cancer. *Biology (Basel)*, *9*(1). doi:10.3390/biology9010004
- Vega-Rubín-de-Celis, S., Zou, Z., Fernández Á, F., Ci, B., Kim, M., Xiao, G., . . . Levine, B. (2018). Increased autophagy blocks HER2-mediated breast tumorigenesis. *Proc Natl Acad Sci U S A*, *115*(16), 4176-4181. doi:10.1073/pnas.1717800115
- Villman, K., Sjöström, J., Heikkilä, R., Hultborn, R., Malmström, P., Bengtsson, N. O., . . . Blomqvist, C. (2006). TOP2A and HER2 gene amplification as predictors of response to anthracycline treatment in breast cancer. *Acta Oncol*, *45*(5), 590-596. doi:10.1080/02841860500543182
- Waks, A. G., & Winer, E. P. (2019). Breast Cancer Treatment: A Review. *Jama*, *321*(3), 288-300. doi:10.1001/jama.2018.19323
- Walerych, D., Napoli, M., Collavin, L., & Del Sal, G. (2012). The rebel angel: mutant p53 as the driving oncogene in breast cancer. *Carcinogenesis*, *33*(11), 2007-2017. doi:10.1093/carcin/bgs232
- Wang, Y., Li, J., Dai, L., Zheng, J., Yi, Z., & Chen, L. (2018). MiR-17-5p may serve as a novel predictor for breast cancer recurrence. *Cancer Biomark*, *22*(4), 721-726. doi:10.3233/cbm-181228

- Wang, Z. (2013). *Androgen-Responsive Genes in Prostate Cancer, Regulation, Function and Clinical Applications* (Z. Wang Ed. 2013 ed.). USA: Springer.
- Wardell, S. E., Marks, J. R., & McDonnell, D. P. (2011). The turnover of estrogen receptor α by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy. *Biochem Pharmacol*, *82*(2), 122-130. doi:10.1016/j.bcp.2011.03.031
- Wee, P., & Wang, Z. (2017). Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)*, *9*(5). doi:10.3390/cancers9050052
- White, E. (2015). The role for autophagy in cancer. *J Clin Invest*, *125*(1), 42-46. doi:10.1172/jci73941
- Wolff, A. C., Hammond, M. E., Schwartz, J. N., Hagerty, K. L., Allred, D. C., Cote, R. J., . . . Hayes, D. F. (2007). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol*, *25*(1), 118-145. doi:10.1200/jco.2006.09.2775
- Wu, X., Ding, M., & Lin, J. (2020). Three-microRNA expression signature predicts survival in triple-negative breast cancer. *Oncol Lett*, *19*(1), 301-308. doi:10.3892/ol.2019.11118
- Wu, Y. M., Su, F., Kalyana-Sundaram, S., Khazanov, N., Ateeq, B., Cao, X., . . . Chinnaiyan, A. M. (2013). Identification of Targetable FGFR Gene Fusions in Diverse Cancers. *Cancer Discov*. doi:10.1158/2159-8290.cd-13-0050
- Xiao, W., Zheng, S., Yang, A., Zhang, X., Zou, Y., Tang, H., & Xie, X. (2018). Breast cancer subtypes and the risk of distant metastasis at initial diagnosis: a population-based study. *Cancer Manag Res*, *10*, 5329-5338. doi:10.2147/cmar.s176763
- Yadav, S., & Couch, F. J. (2019). Germline Genetic Testing for Breast Cancer Risk: The Past, Present, and Future. *Am Soc Clin Oncol Educ Book*, *39*, 61-74. doi:10.1200/edbk_238987
- Yan, L. X., Liu, Y. H., Xiang, J. W., Wu, Q. N., Xu, L. B., Luo, X. L., . . . Chen, J. (2016). PIK3R1 targeting by miR-21 suppresses tumor cell migration and invasion by reducing PI3K/AKT signaling and reversing EMT, and predicts clinical outcome of breast cancer. *Int J Oncol*, *48*(2), 471-484. doi:10.3892/ijo.2015.3287

- Yang, L., Li, Y., Bhattacharya, A., & Zhang, Y. (2019). A recombinant human protein targeting HER2 overcomes drug resistance in HER2-positive breast cancer. *Sci Transl Med*, *11*(476). doi:10.1126/scitranslmed.aav1620
- Yao, J., Weremowicz, S., Feng, B., Gentleman, R. C., Marks, J. R., Gelman, R., . . . Polyak, K. (2006). Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res*, *66*(8), 4065-4078. doi:10.1158/0008-5472.can-05-4083
- Ying, Q., Liang, L., Guo, W., Zha, R., Tian, Q., Huang, S., . . . He, X. (2011). Hypoxia-inducible microRNA-210 augments the metastatic potential of tumor cells by targeting vacuole membrane protein 1 in hepatocellular carcinoma. *Hepatology*, *54*(6), 2064-2075. doi:10.1002/hep.24614
- Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., & Verhaak, R. G. (2015). The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, *34*(37), 4845-4854. doi:10.1038/onc.2014.406
- Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., . . . Song, E. (2007). let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell*, *131*(6), 1109-1123. doi:10.1016/j.cell.2007.10.054
- Yu, X., Liang, J., Xu, J., Li, X., Xing, S., Li, H., . . . Du, H. (2018). Identification and Validation of Circulating MicroRNA Signatures for Breast Cancer Early Detection Based on Large Scale Tissue-Derived Data. *J Breast Cancer*, *21*(4), 363-370. doi:10.4048/jbc.2018.21.e56
- Yu, Z., Wang, C., Wang, M., Li, Z., Casimiro, M. C., Liu, M., . . . Pestell, R. G. (2008). A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J Cell Biol*, *182*(3), 509-517. doi:10.1083/jcb.200801079
- Yun, C. W., & Lee, S. H. (2018). The Roles of Autophagy in Cancer. *Int J Mol Sci*, *19*(11). doi:10.3390/ijms19113466
- Zambrano, J., & Yeh, E. S. (2016). Autophagy and Apoptotic Crosstalk: Mechanism of Therapeutic Resistance in HER2-Positive Breast Cancer. *Breast Cancer (Auckl)*, *10*, 13-23. doi:10.4137/bcbcr.s32791
- Zhang, L., Feizi, N., Chi, C., & Hu, P. (2018). Association Analysis of Somatic Copy Number Alteration Burden With Breast Cancer Survival. *Front Genet*, *9*, 421. doi:10.3389/fgene.2018.00421

- Zheng, L., Chen, L., Zhang, X., Zhan, J., & Chen, J. (2016). TMEM49-related apoptosis and metastasis in ovarian cancer and regulated cell death. *Mol Cell Biochem*, 416(1-2), 1-9. doi:10.1007/s11010-016-2684-3
- Zhu, S., Si, M. L., Wu, H., & Mo, Y. Y. (2007). MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J Biol Chem*, 282(19), 14328-14336. doi:10.1074/jbc.M611393200
- Zhu, S., Wu, H., Wu, F., Nie, D., Sheng, S., & Mo, Y. Y. (2008). MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res*, 18(3), 350-359. doi:10.1038/cr.2008.24
- Zhuyan, J., Chen, M., Zhu, T., Bao, X., Zhen, T., Xing, K., . . . Zhu, S. (2020). Critical steps to tumor metastasis: alterations of tumor microenvironment and extracellular matrix in the formation of pre-metastatic and metastatic niche. *Cell Biosci*, 10, 89. doi:10.1186/s13578-020-00453-9
- Zimmerman, M. W., Liu, Y., He, S., Durbin, A. D., Abraham, B. J., Easton, J., . . . Look, A. T. (2017). c-MYC drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. *Cancer Discov*. doi:10.1158/2159-8290.cd-17-0993

Original publications

Paper I

RESEARCH ARTICLE

High expression of the vacuole membrane protein 1 (*VMP1*) is a potential marker of poor prognosis in HER2 positive breast cancer

Arsalan Amirfallah^{1,2}, Adalgeir Arason^{2,3}, Hjorleifur Einarsson^{1*}, Eydis Thorunn Gudmundsdottir¹, Edda Sigridur Freysteinsdottir³, Kristrun Audur Olafsdottir⁴, Oskar Thor Johannsson⁵, Bjarni Agnar Agnarsson^{4,6}, Rosa Bjork Barkardottir^{2,3}, Inga Reynisdottir^{1,2*}

1 Cell Biology Unit at the Pathology Department, Landspítali—The National University Hospital of Iceland, Reykjavik, Iceland, **2** The Biomedical Center, University of Iceland, Reykjavik, Iceland, **3** Molecular Pathology Unit at the Pathology Department, Landspítali—The National University Hospital of Iceland, Reykjavik, Iceland, **4** Pathology Department, Landspítali—The National University Hospital of Iceland, Reykjavik, Iceland, **5** Department of Oncology, Landspítali—The National University Hospital of Iceland, Reykjavik, Iceland, **6** Faculty of Medicine, University of Iceland, Reykjavik, Iceland

* Current address: Section of Computation and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

* ingar@landspitali.is



OPEN ACCESS

Citation: Amirfallah A, Arason A, Einarsson H, Gudmundsdottir ET, Freysteinsdottir ES, Olafsdottir KA, et al. (2019) High expression of the vacuole membrane protein 1 (*VMP1*) is a potential marker of poor prognosis in HER2 positive breast cancer. PLoS ONE 14(8): e0221413. <https://doi.org/10.1371/journal.pone.0221413>

Editor: Yuan-Soon Ho, Taipei Medical University, TAIWAN

Received: April 9, 2019

Accepted: August 6, 2019

Published: August 23, 2019

Copyright: © 2019 Amirfallah et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Relevant data are within the paper and its Supporting Information files. The paper also includes instructions for accessing data stored in public repositories.

Funding: This work was supported by a grant to IR, RBB, BAA, OTJ and Ad Ar from The Icelandic Centre for Research Fund (152530-051, www.rannis.is), to IR, RBB, BAA and OTJ from The Scientific Fund of Landspítali – The National University Hospital in Iceland (A-2015-039, A-

Abstract

Background

Fusion genes result from genomic structural changes, which can lead to alterations in gene expression that supports tumor development. The aim of the study was to use fusion genes as a tool to identify new breast cancer (BC) genes with a role in BC progression.

Methods

Fusion genes from breast tumors and BC cell lines were collected from publications. RNA-Seq data from tumors and cell lines were retrieved from databanks and analyzed for fusions with SOAPfuse or the analysis was purchased. Fusion genes identified in both tumors ($n = 1724$) and cell lines ($n = 45$) were confirmed by qRT-PCR and sequencing. Their individual genes were ranked by selection criteria that included correlation of their mRNA level with copy number. The expression of the top ranked gene was measured by qRT-PCR in normal tissue and in breast tumors from an exploratory cohort ($n = 141$) and a validation cohort ($n = 277$). Expression levels were correlated with clinical and pathological factors as well as the patients' survival. The results were followed up in BC cohorts from TCGA ($n = 818$) and METABRIC ($n = 2509$).

Results

Vacuole membrane protein 1 (*VMP1*) was the most promising candidate based on specific selection criteria. Its expression was higher in breast tumor tissue than normal tissue ($p = 1 \times 10^{-4}$), and its expression was significantly higher in HER2 positive than HER2 negative breast tumors in all four cohorts analyzed. High expression of *VMP1* associated with breast

2018-034, www.landspitali.is), and grants from Gongum saman (2013 and 2017 to IR and RBB, and 2018 to Ar Am, <http://www.gongumsaman.is/>). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

cancer specific survival (BCSS) in cohort 1 (hazard ratio (HR) = 2.31, CI 1.27–4.18) and METABRIC (HR = 1.26, CI 1.02–1.57), and also after adjusting for HER2 expression in cohort 1 (HR = 2.03, CI 1.10–3.72). BCSS was not significant in cohort 2 or TCGA cohort, which may be due to differences in treatment regimens.

Conclusions

The results suggest that high VMP1 expression is a potential marker of poor prognosis in HER2 positive BC. Further studies are needed to elucidate how VMP1 could affect pathways supportive of tumorigenesis.

Introduction

BC is the most common type of cancer diagnosed in women worldwide [1]. The prognosis and treatment depend on the stage of the disease at diagnosis, the type of tumor, the grade, the proliferation status (Ki67 expression), and the expression of HER2/ERBB2 and the hormonal receptors, estrogen and progesterone receptors. Even though drugs, which are tailored to the genetic make-up of a tumor such as HER2 expression, are increasingly being used, not all tumors respond to treatment and options for further targeted treatment is limited for patients that experience relapse of their disease. Therefore, identifying new genes that support tumor progression in the breast could be used to improve prognosis and follow-up of patients.

Genes that support tumorigenesis most often have undergone changes that result in loss of control or changes in expression patterns. Genetic rearrangements such as amplification, translocations, inversions, insertions and deletions are frequent in breast tumors. Amplified chromosomal regions are well known in breast tumors [2], particularly the amplifications of the *ERBB2* locus at 17q12. It results in the gene's overexpression giving the cell the potential to bypass regulatory mechanisms and support malignant growth. Fusion genes, generated through inter-chromosomal translocations or intrachromosomal changes such as inversions or deletion of chromosomal segments, also can acquire such oncogenic potential [3, 4]. Recurring fusion genes have only been identified in subgroups of breast tumors [5–7] rather than across different subtypes of breast tumors [4, 8]. Most studies have focused on functional chimeric fusion proteins even though they are a minority of fusion genes [4, 8, 9]. Translocations can result in inappropriate expression of genes through promoter switching [10] and loss of 3' UTR regulation by miRNAs [5, 11]. They can activate intragenic miRNAs inappropriately [5] as well as place superenhancers in the vicinity of genes resulting in overexpression of genes in the absence of amplification [12]. As such, a genetic rearrangement that results in a fusion gene may produce a single gene with malignant properties rather than produce a functional chimera made from two genes. Therefore, we postulated that screening fusion genes could be used as a tool to identify potentially novel cancer genes that can affect tumor development. Herein, we describe a screen of fusion genes in a large group of breast tumors and in BC cell lines that identified vacuole membrane protein 1 (*VMP1*) as a gene that may contribute to breast tumor progression.

Materials and methods

Fusion genes from breast tumors and BC cell lines

Fusion genes from breast tumors were collected from three studies [4, 9, 13]. In addition, a list of fusion genes from breast tumors from The Cancer Genome Atlas (TCGA) was purchased

from MediSapiens (www.medisapiens.com). They used the MediSapiens FusionSCOUT pipeline to identify fusion genes in RNA-Seq data. Fusion genes from BC cell lines were collected from publications [14–20]. Furthermore, we analyzed RNA-Seq data from BC cell lines with the fusion finding algorithm SOAPfuse [21]: CAMA-1 (GSM1172856), MDAMB134VI (GSM1172886), MDA-MB-231 (GSM1172889), SUM-225 (GSM1172901), SUM-229 (GSM1172902), SUM52 (GSM1172903), SUM44 (GSM1897347), and UACC893 (GSM1172907/GSM1897353). The paired-end RNA-Seq data from the cell lines were mapped to the human reference genome (hg19) and annotated transcripts (Ensembl release 75) using SOAP2. Then, SOAPfuse was used to identify fusion genes by detecting span and junction reads from the aligned data. Analyses of the RNA-Seq data from the cell lines also were purchased from MediSapiens. Fusion genes in BC cell lines that were identified by both MediSapiens FusionSCOUT pipeline and SOAPfuse were considered for validation.

Cohorts and tissue samples and clinical data

Cohort 1 consisted of 158 BC patients, diagnosed 1987–2003 [22], and cohort 2 consisted of 291 patient, diagnosed 2003–2007 (S1 Table). The relevant patient data were collected from hospital records at Landspítali–The National University Hospital of Iceland as described previously [22]. Primary fresh frozen tumors were obtained from the Department of Pathology as well as six non-neoplastic breast tissue, taken as far away from the tumor as possible. Informed consent was obtained from all patients involved in this study according to the national guidelines. The study was approved by The Icelandic Data Protection Commission (2001/523 and 2002/463) as well as the National Bioethics Committee of Iceland (99/051, 99/051_FS1, VSN-11-105, VSN-15-138). The Nordic cohort consisted of 577 primary breast tumors from patients whose majority was diagnosed 1987–2003 in Finland, Sweden and Iceland (including samples from cohort 1) [23, 24]. TCGA cohort consisted of 818 BC patients diagnosed 1988–2013 [25] and the METABRIC patients were 2,509, diagnosed 1980–2005 [26–28], with data available for both cohorts through cBioPortal [29, 30] and from Rueda et al. [28].

DNA and RNA isolation

DNA and total RNA were extracted from fresh frozen breast tumors from patients in cohort 2 ($n = 291$) and from 6 normal breast tissue samples as well as 1×10^6 MCF-7 cells using Allprep kit DNA/RNA/miRNA (Qiagen no. 80224) according to protocol. The extraction from cohort 1 has been described [22] but in short, total RNA was extracted with Trizol and purified on an RNeasy column according to protocol. The quantity of DNA was measured by Nanodrop 1000 and the RNA quality was measured with Bioanalyzer 2100 RNA 6000 Nano kit (Agilent Technologies, cat. no. 5067–1511) according to protocol. The majority of tumors had $RIN \geq 8$.

Verification of RPS6KB1-VMP1 in MCF-7

MCF-7 was obtained from the American Type Culture Collection. It was cultured in DMEM/F12 (ThermoFisher, cat.no. 11330–032) supplemented with 10% fetal calf serum (ThermoFisher, cat.no. 10270–106), 37°C and 5% CO₂. RNA was extracted as described above and cDNA was synthesized using a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific no. 1622). The RPS6KB1-VMP1 junction was amplified by PCR and then sequenced using primers F: 5′-GAAACTAGTGTGAACAGAGG-3′ and R: 5′-CATAACTTTGTGCCATGGAG-3′.

VMP1 copy number variations

VMP1 copy number data from the Nordic dataset were retrieved from GEO dataset GSE22133 [31] and from the TCGA dataset through cBioPortal [29, 30]. Both sets were measured by comparative genomic hybridization (CGH) on microarrays. The definition of copy number variation (CNV) in the TCGA dataset was used [32].

VMP1 mRNA expression

VMP1 mRNA data for the Nordic dataset were retrieved from GEO (dataset GSE25307) and for the TCGA dataset through cBioPortal [29, 30]. Both sets were measured with gene expression microarrays with probes located at the 3' end of VMP1. Total RNA (0.5 µg) from normal breast tissue and the tumors from cohorts 1 and 2 was used as a template to generate cDNA as described above. Quantification of the VMP1 mRNA level was performed with Taqman Gene Expression Assays spanning exons 10–11 (E10-11; Thermo Fisher Scientific, Taqman/Hs00978589_m1) in both cohorts, and a probe spanning exons 2 and 3 (E2-3; Taqman/Hs00978582_m1) was used to verify the data for cohort 1. TATA-binding protein (TBP, 1702071 Applied Biosystems) was used as a reference gene. All reactions were done in triplicate using 42 cycles with one ng of cDNA as template. VMP1 expression was calculated relative to TBP: $2^{-(\text{mean Ct target} - \text{mean Ct reference})}$. mRNA values were obtained from 141 and 277 tumors in cohorts 1 and 2, respectively. The location of the VMP1 probes is shown in S1 Fig.

Quantification of miR21 expression

cDNA synthesis for miRNA was performed using cDNA synthesis kit II (Exiqon cat. no. 203301) according to the manufacturers protocol. Five ng/µl of RNA from cohort 1 (n = 144) were used. The qRT-PCR reaction was performed with EXIQON primer sets hsa-miR21-5P (YP00204230) and hsa-miR21-3P (YP00204302) along with ExiLent SYBR Green master mix and hsa-miR16-5P (YP00205702) as reference gene. All reactions were done in triplicate using 40 cycles.

Statistical analysis

The statistical program R version 3.4.3 was used [33]. The microarray DNA and mRNA measurements from the Nordic dataset as well as the DNA, mRNA and miRNA measurements from cohorts 1 and 2 were transformed with log₂ to normalize the data. The mRNA values from the METABRIC and TCGA cohorts, available from cBioPortal, are Z-scores. Co-amplification of *ERBB2* and *VMP1* DNA levels was analyzed with χ^2 -test. Correlation between DNA and mRNA levels, or mRNA and miRNA expression, was performed by calculating the Pearson correlation coefficient using normalized values. The correlation analyses between mRNA levels and the clinicopathological characteristics were performed with Student's t-test or ANOVA. P-values below 0.05 were considered significant.

The Kaplan-Meier and log rank test were used to estimate survival using the survival package and the survminer package in R. Survival analysis was based on tumor VMP1 mRNA levels measured by microarrays in the Nordic (n = 553), TCGA (n = 421), and METABRIC (n = 1904) cohorts, and by qPCR with probe E10-11 in cohorts 1 (n = 141) and 2 (n = 277). The tumors were classified as expressing high VMP1 mRNA (\geq mean + 1 SD) or normal VMP1 mRNA ($<$ mean + 1 SD). Hazard ratio (HR) calculation based on VMP1 mRNA levels and clinicopathological characteristics was performed with Cox regression analysis [34]. Due to missing data for VMP1 mRNA as well as lack of complete clinical data in some cohorts the numbers of patient samples in the analyses are lower than the actual number of patients.

Results

A screen of fusion genes identifies VMP1 as a candidate

The generation of fusion genes may lead to loss of control and affect expression of the gene partners. We wanted to explore whether the genes that constitute fusion genes could be used to detect a gene that supports breast cancer development. Therefore, a screen of fusion genes was performed. It entailed the comparison of fusion genes, identified in breast tumors, with fusion genes identified in breast tumor cell lines. Cell lines tend to be aggressive and we reasoned that studying them would increase the likelihood of detecting a gene which is significant in the progression of BC. Fusion genes from BC cell lines were collected from publications [14–18, 20, 35] and RNA-Seq data were analyzed by fusion finding algorithms [21] as described in methods. Information regarding fusion genes and potential fusion genes from breast tumors were acquired from three studies [4, 9, 13] or from MediSapiens. In all, 183 fusion genes (paired genes) were acquired from 45 BC cell lines while 5319 fusion genes were acquired from 1724 breast tumors. The tumors and the cell lines had 15 fusion genes in common. They had to meet the following criteria to merit further analyses: 1) have a similar breakpoint in breast tumors and cell lines, 2) be recurrent in tumors, 3) not be located within an amplicon carrying a known oncogene unless it was part of the fusion, and 4) possess a function supportive of tumorigenesis (available through publications). Five fusion genes met these criteria (Table 1). They were all verified by PCR-amplification and sequencing in their respective cell lines (S2 Fig).

To distinguish which of the 10 genes that constituted the five fusions could be of consequence in BC progression, the copy number of the genes was analyzed and correlated with the respective mRNA levels. CGH microarray and gene expression data from our earlier study on 577 Nordic tumors [24] were used as well as data retrieved from TCGA [25]. Genes that are amplified and with highly correlating gene expression can signify an oncogene. Correlation between DNA and mRNA was highest for *CCDC6* ($r = 0.66$), *GATAD2B* ($r = 0.54$), *RPS6KB1* ($r = 0.83$) and *VMP1* ($r = 0.70$) in the cohort from TCGA (S2 Table). *CCDC6* was not amplified and even though *GATAD2B* was amplified in the TCGA cohort it was not amplified in the Nordic cohort. *RPS6KB1* and *VMP1* were the genes most frequently amplified in both cohorts, with amplification close to 11% in tumors from TCGA (S2 Table). *RPS6KB1* and *VMP1* are adjacent genes at 17q23. A tandem duplication of the locus was found in MCF-7 cells that resulted in a fusion between *RPS6KB1* and *VMP1* [36]. Although the tandem duplication was not common, they observed the *RPS6KB1-VMP1* fusion transcript, with varying breakpoints,

Table 1. Five fusion genes in common between breast tumors and breast cancer cell lines.

5' fusion gene partner	3' fusion gene partner	No of fusions (%)	Cell lines
CCDC6	ANK3	2 (0.12)	UACC893
ESR1	CCDC170	11 (0.64)	ZR751
GATAD2B	NUP210L	1 (0.06)	MCF-7
ITGB6	RBMS1	1 (0.06)	UACC893
RPS6KB1	VMP1	5 (0.29)	MCF-7

Fusion genes were analyzed in a total of 1724 breast tumors. The number of breast tumors carrying the fusion genes that were found in common between breast tumors and breast cell lines is shown in this table. The common fusions were analyzed in other tumor types through this website: www.tumorfusions.org [4]. *GATAD2B-NUP210L* appeared once in these tumor types: uterine carcinosarcoma (UCS), lung adenocarcinoma (LUAD) and ovarian tumors (OV). *ITGB6-RBMS1* was found in two tumors from bladder cancer (BLCA).

<https://doi.org/10.1371/journal.pone.0221413.t001>

in 22 tumors from a cohort of 70 BC patients from Singapore [36]. The fusion was observed in only five of 1724 tumors in our study (Table 1), only one of which was HER2 positive. The discrepancy in the frequency could be due to the ethnicity of the patients, from Singapore [36] as opposed to cohorts in which the majority of patients were of European descent [8, 26, 27], or it could be due to the method, specific screening for the *RPS6KB1-VMP1* fusion [36] as opposed to searching for fusion genes using RNA-Seq data, which was the basis of our study. The *RPS6KB1-VMP1* fusion was not enriched in HER2 positive tumors in the data that we used (one in five tumors) and of the 45 cell lines that we used it was only found in MCF-7, which is HER2 negative. Again, the depth of RNA sequencing and different fusion finding algorithms used in the various studies may be the reason. Interestingly, *VMP1* was found as a 3' partner in fusion transcripts in 16 of the 1724 tumors (0.93%). Four of the tumors with *VMP1* fusions, or 25%, were HER2 positive while 16% of the tumors with non-*VMP1* fusions were HER2 positive. This is in accordance with Persson et al. [5], who showed *VMP1* fusion transcripts to be enriched among HER2 positive tumors. Interestingly, the majority of the in frame fusion transcripts identified by Inaki et al. [36] included only the first exon of *RPS6KB1* and the C-terminal half of *VMP1*. Thus, the functional activity of the chimeric protein would be expected to stem from *VMP1*. In addition, *RPS6KB1* has been shown to associate with HER2 positivity and a worse outcome in BC ([37] and references therein). Thus, *VMP1* was selected as a candidate.

VMP1 is a potential player in breast tumorigenesis

Further analyses were performed in the Nordic and TCGA cohorts to explore the potential role of *VMP1* in breast tumor development. The highest correlation between *VMP1* DNA and mRNA was observed in tumors with *VMP1* amplification (TCGA: $r = 0.72$, $p = 3.4 \times 10^{-9}$) and in tumors with overexpression of *ERBB2*, either according to molecular subtype [38] (Nordic cohort: $r = 0.79$, $r = 4.31 \times 10^{-8}$) or HER2 expression (TCGA: $r = 0.82$, $p = 2.2 \times 10^{-16}$). To examine whether there were consequences of high *VMP1* expression, survival analyses were performed in the Nordic and TCGA cohorts. They suggested shorter overall survival (OS) in BC patients carrying tumors with high levels of *VMP1* mRNA (TCGA: log rank p -value = 0.023 and Nordic: log rank $p = 0.064$, Fig 1). The hazard ratio (HR) was 2.10 (CI 1.09–4.04) in TCGA and 1.37 (CI 0.98–1.91) in the Nordic cohort. One indication of oncogenic properties of a gene is higher expression levels in tumors than normal tissue. To compare expression in our cohorts, RNA was extracted from tumors in cohorts 1 ($n = 141$) and 2 ($n = 277$), and from the available normal breast tissue samples ($n = 6$) from cohort 2. *VMP1* mRNA was measured by qPCR. It was found to be significantly higher in breast tumors from cohort 1 ($p = 1 \times 10^{-4}$) and cohort 2 ($p = 3 \times 10^{-4}$) than in normal breast tissue (S3 Fig).

VMP1 is located at 17q23, a chromosomal region whose copy number is increased in up to 22% of primary breast tumors depending on their histological origin [39]. Many genes reside within the amplified region but *RPS6KB1*, *MIR21* [40], and *PPMD1* [41] have been suggested as drivers of the amplification with oncogenic properties. In the TCGA cohort, *VMP1* mRNA positively correlated with the mRNAs of *RPS6KB1* ($r = 0.67$, $p < 2.2 \times 10^{-16}$) and *PPMD1* ($r = 0.58$, $p < 2.2 \times 10^{-16}$), and with *miR21* ($r = 0.50$, $p = 1.75 \times 10^{-12}$). As expression from these genes could affect survival on their own and thus confound the effect observed with *VMP1*, a Cox regression analysis was performed to adjust for their expression. Expression of these genes did not attenuate the effect of *VMP1* mRNA on OS in the TCGA cohort (S3 Table).

MIR21 overlaps the 3' end of *VMP1* [42, 43], and many of the fusion gene breakpoints within *VMP1* occur just prior to *MIR21* (<http://www.tumorfusions.org/>, [5, 36]). The probes used to measure *VMP1* mRNA in TCGA and cohorts 1 and 2 (spanning E10–11) were located

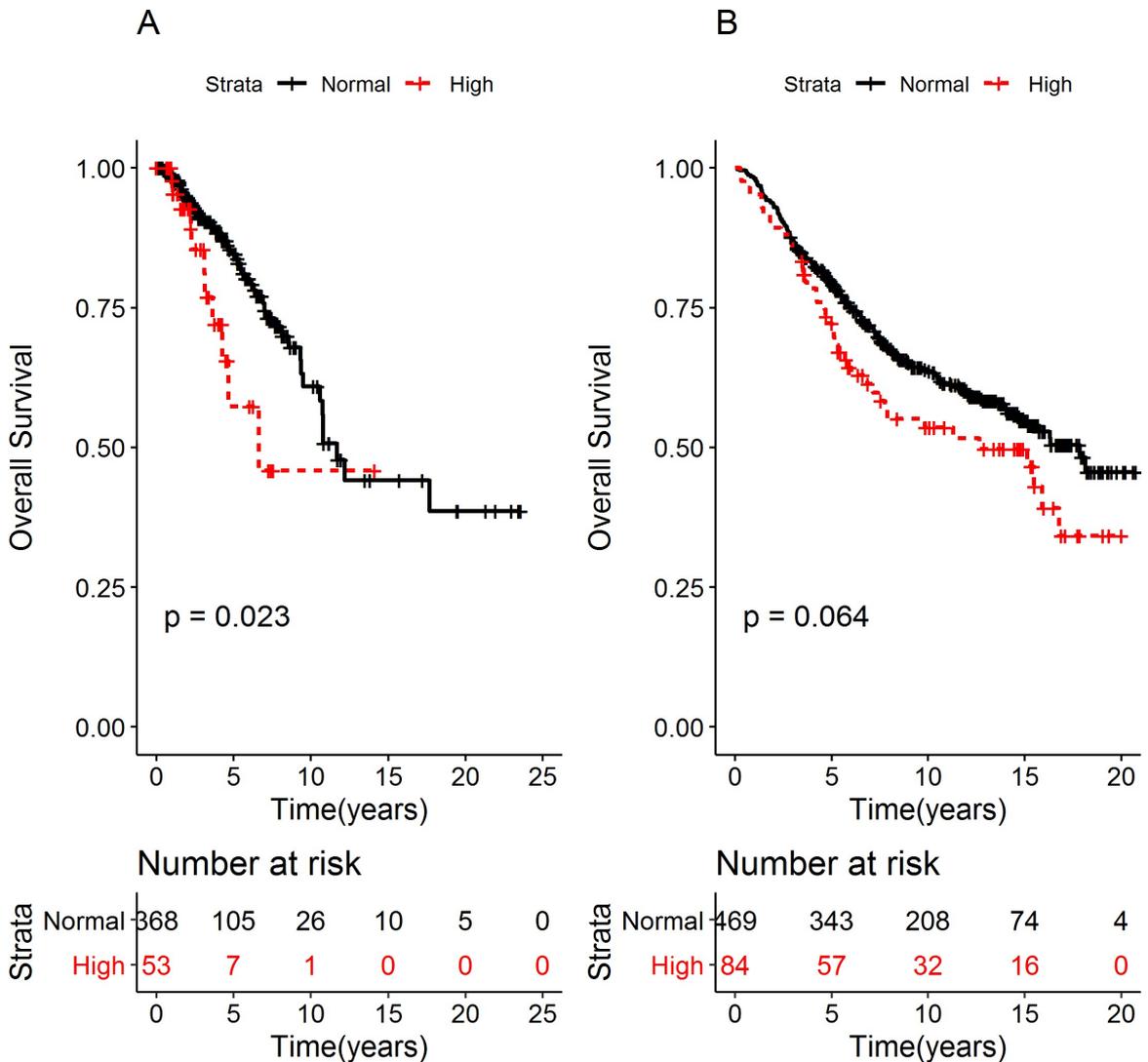


Fig 1. High VMP1 mRNA expression is correlated with shorter OS. Overall survival (OS) was examined in breast cancer patients in (A) TCGA and (B) the Nordic cohort. The patients were divided into two groups according to VMP1 mRNA levels: tumors expressing high VMP1 mRNA (high \geq mean + 1 SD) and normal VMP1 (normal < mean + 1 SD). The log rank p-values are indicated in the graphs. The number of patients at risk is shown below the graphs in tables at the indicated time points. The median OS for the TCGA cohort was 11.68 and 6.62 years for patients expressing normal and high levels of VMP1 mRNA, respectively. The hazard ratio (HR) for OS was 2.10, CI 1.09–4.04. The median OS in the Nordic cohort was 16.3 and 12.6 years for patients expressing normal and high VMP1 mRNA, respectively. The HR was 1.37, CI: 0.98–1.91.

<https://doi.org/10.1371/journal.pone.0221413.g001>

in the C-terminus and potentially can detect pri-miRNA-21. Thus, VMP1 mRNA was measured with a probe spanning E2-3 in cohort 1, and the expression of the mature miRNA products, hsa-mir-21-5p and hsa-mir-21-3p was measured as well. The correlation between the VMP1 E2-3 and E10-11 probes was high ($r = 0.85$, $p < 0.001$). The VMP1 mRNA probes did

not correlate with hsa-mir-21-5p or hsa-mir-21-3p ($p > 0.05$), indicating that the signal from the E10-11 probe reflected VMP1 mRNA levels.

Taking the data together, they suggest that VMP1 may have oncogenic properties, and we wanted to explore whether VMP1 mRNA levels could have a prognostic value.

VMP1 mRNA level is high in breast tumors that express HER2

In order to understand whether VMP1's expression levels could indicate severity of disease the mRNA values were correlated with the tumor's clinical and pathological characteristics. In breast tumors from cohort 1, higher VMP1 expression level was observed in ERBB2/HER2 positive tumors based on classification with immunohistochemistry (HER2, $p = 7 \times 10^{-4}$) or molecular subtyping ($p = 5 \times 10^{-6}$, Fig 2 and Table 2, [38]). There was a highly significant association between HER2 positivity and increased VMP1 mRNA levels in all cohorts: in cohort 2 $p = 0.004$ (S4 Table), in TCGA $p = 0.003$ (S5 Table), and in METABRIC $p < 2.2 \times 10^{-16}$ (S6 Table). The significant association between VMP1 mRNA levels and the intrinsic subtypes in TCGA and METABRIC were due to high levels of VMP1 mRNA in ERBB2 and luminal B subtypes, which include HER2 positive tumors, and low levels of VMP1 mRNA in the basal subtype, which reflected low VMP1 expression in ER negative tumors (TCGA: $p = 7 \times 10^{-6}$ and METABRIC: $p = 0.01$). This result was supported at the genomic level since the loci hosting *ERBB2* (17q12) and *VMP1* (17q23) were frequently co-amplified as has been published [23, 44] and seen in the Nordic cohort (χ^2 test $< 2.2 \times 10^{-16}$). The linear correlation between the CNVs of *ERBB2* and *VMP1* was low (CNV $r = 0.28$, $p = 4.6 \times 10^{-7}$) as well as between their mRNA ($r = 0.30$, $p = 7.81 \times 10^{-14}$) indicating that VMP1 expression was not high in all *ERBB2* amplified or highly expressing tumors. The data show that VMP1 is highly expressed or amplified in some HER2 positive or *ERBB2* amplified tumors, which may indicate a potential interaction between the two genes.

High VMP1 mRNA is associated with shorter survival

To analyze whether VMP1 mRNA status could predict the outcome of BC patients, survival analyses were performed. Breast cancer specific survival (BCSS) was used rather than OS, which may be due to other diseases in addition to BC. Cohort 1 BC patients with tumors expressing high VMP1 mRNA level had shorter BCSS than patients with normal level VMP1 mRNA (log rank $p = 0.0045$, Fig 3A). The median time of BCSS was 3.75 years for high and 13.22 years for normal VMP1 mRNA, respectively. The HR was 2.31 (CI 1.27–4.18). In the METABRIC cohort high VMP1 mRNA associated with shorter BCSS (log rank $p = 0.032$) with median survival at 21.7 years with high VMP1 versus 23.5 years for normal VMP1 (Fig 3D). The HR was 1.26 (CI 1.02–1.57). There was not an association between high VMP1 mRNA and BCSS in cohort 2 (log rank $p = 0.49$ Fig 3B) or in the cohort from TCGA (log rank $p = 0.12$, Fig 3C). Because VMP1 is necessary for the initial steps of autophagy [45] and autophagy is high in metastatic tumors [46], the effect of high VMP1 levels on distant recurrence free survival (DRFS) was analyzed in the two cohorts for which there were data. High VMP1 was significantly associated with shorter DRFS in cohort 1 (log rank $p = 0.0017$; HR = 2.54, CI 1.39–4.66) (Fig 4A) as well as METABRIC (log rank $p = 0.041$; HR = 1.26, CI 1.00–1.57) (Fig 4B). Since HER2 is a potent oncogene and VMP1 was most highly expressed in HER2 positive tumors, the possibility remained that HER2 could be confounding the effect of high VMP1 on survival. Taking into account the effect of HER2 on DRFS revealed that in cohort 1 HR was reduced to 1.95 but it was still significant (CI 1.04–3.68) whereas in METABRIC the HR was no longer significant (HR 1.06, CI 0.84–1.34). This suggests that VMP1 mediates some of the effect on survival in cohort 1 but in METABRIC it was due to HER2. It would be ideal to

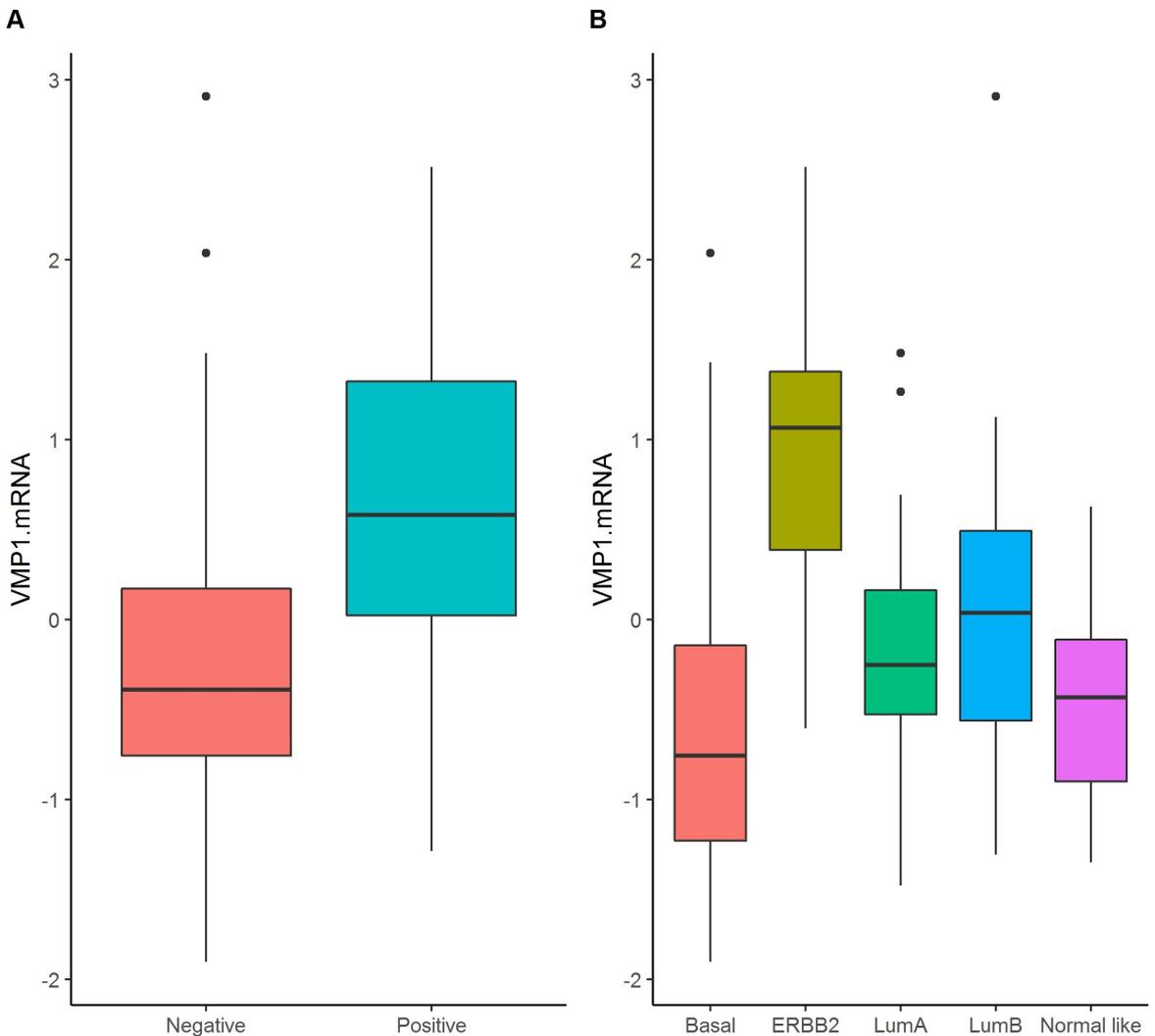


Fig 2. VMP1 mRNA is higher in HER2 positive and ERBB2 breast tumor subtype. VMP1 mRNA was examined according to (A) HER2 expression and (B) molecular subtype in breast tumors from cohort 1. VMP1 mRNA levels were compared between HER2 positive ($n = 23$) and HER2 negative ($n = 117$) tumors, and according to the molecular subtypes (basal = 24, ERBB2 = 14, luminal A = 43, luminal B = 30, normal-like = 12). The p-value was calculated with values normalized by log2 using a t-test for HER2 expression and ANOVA for the molecular subtypes.

<https://doi.org/10.1371/journal.pone.0221413.g002>

analyze the association of VMP1 with survival in a large HER2 positive cohort that has not received trastuzumab or another treatment directed against HER2. The patients in cohort 1 and METABRIC did not receive trastuzumab but only the METABRIC dataset had enough tumors to attempt an analysis of BCSS and DRFS in HER2 positive tumors ($n = 220$). In the METABRIC/HER2 positive cohort high VMP1 was not significantly associated with shorter BCSS (35 vs 185 tumors with high versus low VMP1 mRNA, log rank $p = 0.29$) but the

Table 2. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in cohort 1.

Characteristic	n = 141	VMP1 mRNA level median (25 th , 75 th)	p-value
Age			
≥ 50	85	-0.23 (-0.59, 0.28)	0.67
< 50	56	-0.28 (-0.78, 0.57)	
Estrogen receptor			
positive	90	-0.20 (-0.59, 0.32)	0.41
negative	47	-0.39 (-1.03, 0.50)	
unknown	4		
Progesterone receptor			
positive	70	-0.19 (-0.57, 0.43)	0.42
negative	64	-0.31 (-0.97, 0.37)	
unknown	7		
HER2 status			
positive	23	0.58 (0.02, 1.26)	7x10 ^{-4*}
negative	117	-0.36 (-0.75, 0.18)	
unknown	1		
Receptors ER and HER2			
ER ⁻ and HER2 ⁻	32	-0.56 (-1.14, -0.10)	6.4x10 ^{-6*}
ER ⁻ and HER2 ⁺	14	0.51 (-0.13, 1.20)	
ER ⁺ and HER2 ⁻	82	-0.24 (-0.59, 0.23)	
ER ⁺ and HER2 ⁺	8	0.48 (0.29, 1.25)	
unknown	5		
Tumor size (mm)			
> 20	97	-0.31 (-0.69, 0.40)	0.90
≤ 20	44	-0.14 (-0.62, 0.31)	
Histological type			
IDC ^a	121	-0.14 (-0.65, -0.43)	0.24
ILC ^b	12	-0.35 (-0.56, -0.13)	
other	8	-0.37 (-0.74, -0.10)	
Nodal status			
positive	72	-0.20 (-0.55, 0.57)	0.16
negative	55	-0.13 (-0.88, 0.38)	
unknown	14		
Ki 67			
High	41	-0.46 (-0.85, 0.28)	0.15
Low	97	-0.17 (-0.53, 0.51)	
Unknown	3		
Histological grade			
1	12	-0.49 (-0.67, 0.26)	0.08
2	80	-0.15 (-0.52, 0.25)	
3	48	-0.25 (-0.80, 0.78)	
unknown	1		
Metastasis			
Positive	59	-0.14 (-0.52, 0.60)	0.03*
Negative	81	-0.36 (-0.75, 0.25)	
unknown	1		
Intrinsic subtype			
Basal	24	-0.75 (-1.22, -0.14)	5x10 ^{-6*}

(Continued)

Table 2. (Continued)

Characteristic	n = 141	VMP1 mRNA level median (25 th , 75 th)	p-value
ERBB2	14	0.82 (0.37, 1.35)	
Luminal A	43	-0.24 (-0.51, 0.17)	
Luminal B	30	0.03 (-0.56, 0.49)	
Normal-like	12	-0.47 (-0.87, -0.17)	
unknown	18		
Familial status			
BRCA2	27	-0.43 (-0.89, 0.39)	0.31
Non-BRCA2	114	-0.20 (-0.64, 0.38)	

The table shows the median and the 25th and 75th percentiles. One tumor was BRCA1 positive and it was not used in the familial status calculations. The p-value is calculated with log₂ transformed data using a t-test or ANOVA.

*Significant difference $p < 0.05$.

^aIDC: Invasive ductal tumors.

^bILC: Invasive lobular tumors.

<https://doi.org/10.1371/journal.pone.0221413.t002>

association was suggestive when analyzed for DRFS (log rank $p = 0.085$) (S4 Fig). Even though HER2 is a confounder, there appears to be an effect on survival by VMP1 albeit weak (cohort 1). VMP1 can be activated by HER2 through the PI3K/AKT pathway via GLI3-p300 [47] and independent of HER2 e.g. by the hypoxia induced factor HIF1 α [48]. Further analyses in cell based systems are necessary to understand how VMP1 contributes to BC progression.

Discussion

This study describes how fusion genes were used as a tool to identify potential new BC genes with a role in breast tumor development. VMP1 was the strongest candidate based on our selection criteria. VMP1 mRNA was most highly expressed in HER2 positive tumors, and the results suggest that high VMP1 mRNA may signal worse prognosis for BC patients, most likely in those with HER2 positive tumors.

VMP1 mRNA levels were higher in HER2 positive tumors than HER2 negative tumors in all cohorts analyzed. However, high VMP1 mRNA levels were associated with shorter BCSS in cohort 1 and METABRIC but not in cohort 2 and TCGA. The discrepancy may be due to the extended period of tumor collection and as a result different treatments and the introduction of new drugs during the period. It is tempting to speculate that the differences in the survival analyses between cohorts hinges on trastuzumab because diagnoses of the patients in cohort 1, 1987–2003, preceded the approval of trastuzumab in Iceland, while the diagnoses in cohort 2, 2003–2007, succeeded it, and VMP1 only associated with survival in the former cohort. Furthermore, the patients in the METABRIC cohort, where high VMP1 level associated with shorter BCSS, did not receive trastuzumab [26]. The BC patients in the TCGA cohort were diagnosed over an extended period, 1987–2013, that probably included different types of treatment. However, when HER2 expression was taken into account, the effect on survival was only significant in cohort 1 but not in METABRIC. This was also true when DRFS was analyzed. The discrepancy could be due to different treatments that the patients in these cohorts received, treatments that can have confounding effects on the survival analyses. The four cohorts were diagnosed over extended time periods and as a result they obtained varied drug treatments. Adjusting for all of them in a survival analysis is complex because a subgroup of patients could be selected inadvertently. E.g., in one of the cohorts the patients who were lymph node negative with ER positive tumors received no chemotherapy while the lymph

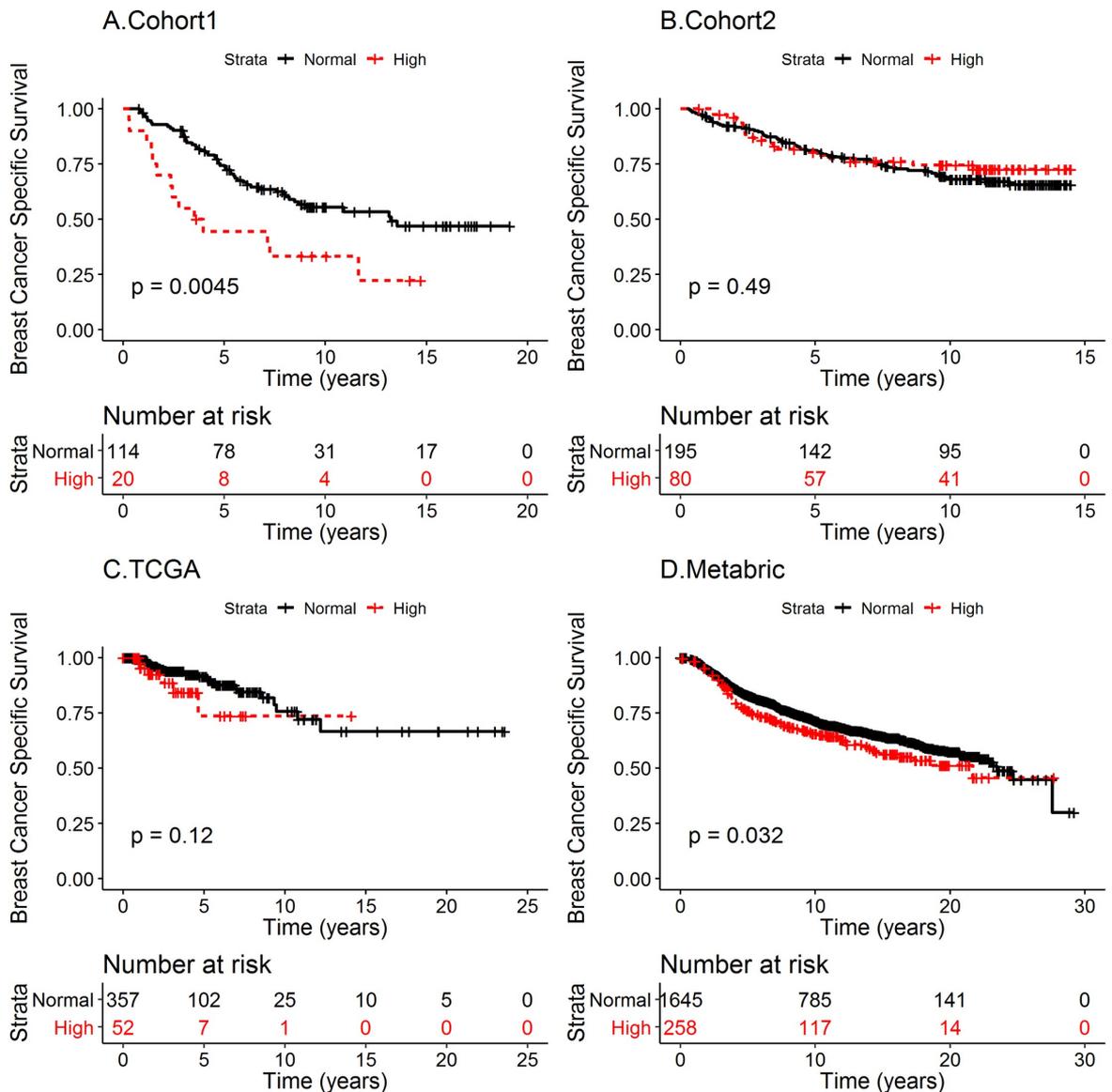


Fig 3. High VMP1 mRNA expression is correlated with shorter BCSS in cohort 1 and METABRIC. Breast cancer specific survival (BCSS) was analyzed in (A) cohort 1, (B) cohort 2, (C) TCGA and (D) METABRIC. The patients were divided into two groups according to VMP1 mRNA levels: tumors expressing high VMP1 mRNA (high \geq mean + 1 SD) and normal VMP1 (normal < mean + 1 SD). The log rank p-values are indicated in the figures. The number of patients at risk is shown below the graphs at the indicated timepoints. The median BCSS was 13.22 and 3.75 years for patients expressing normal and high VMP1 mRNA, respectively, in cohort 1, and 23.5 and 21.7 years in the METABRIC cohort. The hazard ratio (HR) for BCSS in cohort 1 was 2.31 (CI 1.27–4.18), and after adjusting for HER2 expression the HR was 2.03 (CI 1.00–3.72). In METABRIC HR was 1.26 (CI 1.02–1.57) and after adjusting for HER2 expression it was HR = 1.03 (CI 0.82–1.30).

<https://doi.org/10.1371/journal.pone.0221413.g003>

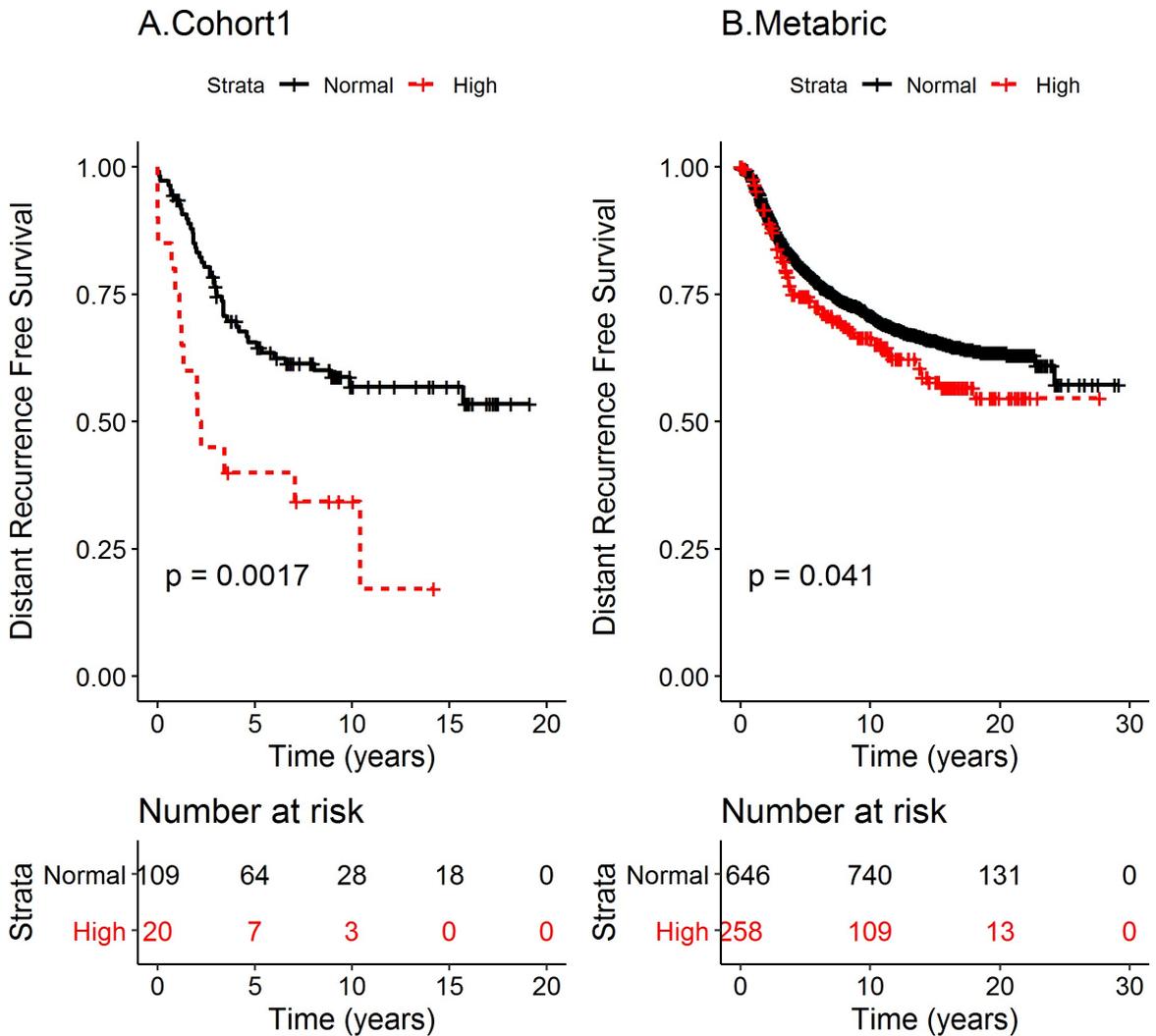


Fig 4. High VMP1 mRNA expression is correlated with shorter DRFS in cohort 1 and METABRIC. Distant recurrence free survival (DRFS) was analyzed in (A) cohort 1 and (B) METABRIC. The patients were divided into two groups according to VMP1 mRNA levels: tumors expressing high VMP1 mRNA (high \geq mean + 1 SD) and normal VMP1 (normal < mean + 1 SD). The log rank p-values are indicated in the figures. The hazard ratio (HR) for DRFS in cohort 1 was 2.54, CI (1.39–4.66), and after adjusting for HER2 expression the HR was 1.95 (CI 1.04–3.68). In METABRIC the HR was 1.26 (CI 1.00–1.57) and after adjusting for HER2 expression it was HR = 1.06 (CI 0.84–1.34).

<https://doi.org/10.1371/journal.pone.0221413.g004>

node positive patients with ER negative tumors received chemotherapy. Also, breaking the cohorts down according to drug treatments would reduce the numbers in the cohorts resulting in less power. Therefore, we did not include drug treatments in the analyses. However, they may explain the different results in the cohorts. Also, we performed tumor microarrays to determine whether the VMP1 mRNA levels reflected the protein levels, but correlation of

mRNA and protein expression could not be assessed due to background staining with the anti-VMP1 antibody.

VMP1 is a transmembrane protein that is associated with the endoplasmic reticulum, Golgi and intracellular vesicles [49]. VMP1 is important for cellular membrane biology as lack of the protein results in defects in endosome trafficking and Golgi morphology [50]. It also has a role in cell adhesion [51], early autophagosome formation [45] and it controls contact between the endoplasmic reticulum and the isolation membranes that precede the formation of the autophagosome [52]. Domains within the protein appear to be highly conserved between species, even bacteria [53]. VMP1's role in BC is not well known. In ovarian tumors VMP1 has been shown to be highly expressed promoting proliferation and metastasis [49] while in colorectal and hepatocellular cancer cells high levels of VMP1 decrease proliferation, invasion and metastasis [54, 55]. This discordance could be due to tumor type but it may also be due to VMP1's role in autophagy. Autophagy has been suggested to act as a tumor suppressor or tumor promoter depending on context [56], and its activity fluctuates during tumor development [46]. VMP1 interacts with the autophagy regulator BECN1 [57], whose interaction with HER2 inhibits autophagy [58]. In pancreatic cells, VMP1-induces autophagy and the KRAS^{G12D} mutation co-operates to promote the formation of pancreatic ductal adenocarcinoma [59]. Hypoxia inducible factors (HIFs) are activated in regions of rapidly growing tumors that are often poorly oxygenated. HIF1 α expression increases VMP1-induced autophagy that results in less cell death in response to photodynamic therapy [48]. HER2 uses the hypoxia system as it regulates HIF2 α under normoxic and hypoxic conditions to upregulate hypoxia genes that help the tumor to survive [60]. Thus, VMP1 could support tumor progression at various points, both independent and dependent on HER2.

ERBB2, at 17q12, is amplified in 15% of breast tumors [61]. Genes that are co-amplified with ERBB2 can result in resistance to anti-HER2 therapy. Genes within the 17q12-21 locus are frequently co-amplified with ERBB2, and some of them, like GRB7, have been shown to co-operate with HER2 [62]. Amplification or expression of *TOP2A* is an indicator of worse prognosis in BC patients [63]. It expresses topoisomerase II α , which is a target of anthracycline. *TOP2A* has been suggested as a biomarker for treatment in HER2 positive BC even though further research is needed [64]. 17q23, where *VMP1* resides, is amplified in 20% of *ERBB2* amplified tumors [23, 40, 44]. A recent study demonstrated that overexpression of only PPM1D or miR21 from the 17q23 locus co-operated with HER2 to induce growth in soft agar in murine mammary tumor virus cells expressing HER2 (MMTV-ErbB2) [65]. In addition these genes increased resistance to therapy targeting HER2 but targeting HER2 and PPM1D and/or miR21 reduced the tumor burden of the cells. Neither PPM1D nor miR21 abolished the effect of high VMP1 mRNA on survival. The induction of autophagy by VMP1 may be important for the development of drug resistance [66]. Chemotherapy can trigger autophagy [67], which has been shown to contribute to the development of resistance to drugs, including HER2 blockers [68, 69] and tamoxifen [70, 71]. Thus, identifying genes that induce resistance in HER2 positive tumors can benefit patients in the form of additional therapies.

Conclusions

Taken together, the data presented suggest that high VMP1 expression may be a marker of poor prognosis in BC, particularly in HER2 positive breast tumors. Since VMP1 is important for autophagosome formation, HER2 positive tumors with high VMP1 may more readily initiate autophagy, which provides building blocks for replication and survival, and therefore the patients could be more prone to relapse. Further studies are needed in cell based systems to elucidate the role of VMP1 in breast tumor development.

Supporting information

S1 Fig. Location of the probes within VMP1.
(PDF)

S2 Fig. The sequenced junction of fusion gene *RPS6KB1-VMP1*.
(PDF)

S3 Fig. VMP1 mRNA is higher in breast tumors than normal breast tissue in cohorts 1 and 2.
(PDF)

S4 Fig. A suggestive association was observed between high VMP1 mRNA and DRFS in METABRIC/HER2 positive patients.
(PDF)

S1 Table. Patient characteristics of cohort 2.
(PDF)

S2 Table. Amplification and correlation between DNA and mRNA of the gene partners that constitute the fusion genes.
(PDF)

S3 Table. The effect of VMP1 on overall survival in TCGA was not attenuated by *RPS6KB1*, *PPM1D* and *miR21*.
(PDF)

S4 Table. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in cohort 2.
(PDF)

S5 Table. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in TCGA.
(PDF)

S6 Table. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in METABRIC.
(PDF)

Acknowledgments

The authors would like to thank Dr. Helga M. Ogmundsdóttir for her thorough review of the manuscript and Dr. Katrin Halldorsdottir for her assistance with figures. The authors also thank the ladies in Gongum saman for their enthusiasm as well as their efforts to support scientists working on breast cancer in Iceland.

Author Contributions

Conceptualization: Rosa Bjork Barkardottir, Inga Reynisdottir.

Data curation: Adalgeir Arason.

Formal analysis: Arsalan Amirfallah, Hjorleifur Einarsson, Inga Reynisdottir.

Funding acquisition: Rosa Bjork Barkardottir, Inga Reynisdottir.

Investigation: Arsalan Amirfallah, Hjorleifur Einarsson.

Methodology: Rosa Bjork Barkardottir, Inga Reynisdottir.

Project administration: Rosa Bjork Barkardottir, Inga Reynisdottir.

Resources: Adalgeir Arason, Oskar Thor Johannsson, Bjarni Agnar Agnarsson, Rosa Bjork Barkardottir.

Supervision: Inga Reynisdottir.

Validation: Arsalan Amirfallah, Inga Reynisdottir.

Visualization: Arsalan Amirfallah.

Writing – original draft: Inga Reynisdottir.

Writing – review & editing: Arsalan Amirfallah, Adalgeir Arason, Hjorleifur Einarsson, Eydis Thorunn Gudmundsdottir, Edda Sigridur Freysteinsdottir, Kristrun Audur Olafsdottir, Oskar Thor Johannsson, Bjarni Agnar Agnarsson, Rosa Bjork Barkardottir, Inga Reynisdottir.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: a cancer journal for clinicians*. 2018; 68(1):7–30.
2. Letessier A, Sircoulomb F, Ginestier C, Cervera N, Monville F, Gelsi-Boyer V, et al. Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers. *BMC Cancer*. 2006; 6:245. <https://doi.org/10.1186/1471-2407-6-245> PMID: 17040570
3. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet*. 2013; 14(10):703–18. <https://doi.org/10.1038/nrg3539> PMID: 24022702
4. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*. 2015; 34(37):4845–54. <https://doi.org/10.1038/onc.2014.406> PMID: 25500544
5. Persson H, Søkilde R, Häkkinen J, Pirona AC, Vallon-Christersson J, Kvist A, et al. Frequent miRNA-convergent fusion gene events in breast cancer. *Nat Commun*. 2017; 8(1):788. <https://doi.org/10.1038/s41467-017-01176-1> PMID: 28983113
6. Tognon C, Knezevich SR, Huntsman D, Roskelley CD, Melnyk N, Mathers JA, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell*. 2002; 2(5):367–76. PMID: 12450792
7. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012; 486(7403):405–9. <https://doi.org/10.1038/nature11154> PMID: 22722202
8. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016; 534(7605):47–54. <https://doi.org/10.1038/nature17676> PMID: 27135926
9. Asmann YW, Necela BM, Kalari KR, Hossain A, Baker TR, Carr JM, et al. Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res*. 2012; 72(8):1921–8. <https://doi.org/10.1158/0008-5472.CAN-11-3142> PMID: 22496456
10. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005; 310(5748):644–8. <https://doi.org/10.1126/science.1117679> PMID: 16254181
11. Parker BC, Annala MJ, Cogdell DE, Granberg KJ, Sun Y, Ji P, et al. The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. *J Clin Invest*. 2013; 123(2):855–65. <https://doi.org/10.1172/JCI67144> PMID: 23298836
12. Zimmerman MW, Liu Y, He S, Durbin AD, Abraham BJ, Easton J, et al. c-MYC drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. *Cancer Discov*. 2017.
13. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016; 534(7605):47–54. <https://doi.org/10.1038/nature17676> PMID: 27135926

14. Asmann YW, Hossain A, Necela BM, Middha S, Kalari KR, Sun Z, et al. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res.* 2011; 39(15):e100. <https://doi.org/10.1093/nar/gkr362> PMID: 21622959
15. Kalyana-Sundaram S, Shankar S, Deroo S, Iyer MK, Palanisamy N, Chinnaiyan AM, et al. Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia.* 2012; 14(8):702–8. <https://doi.org/10.1593/neo.12914> PMID: 22952423
16. Kangaspeska S, Hultsch S, Edgren H, Nicorici D, Murumägi A, Kallioniemi O. Reanalysis of RNA-sequencing data reveals several additional fusion genes with multiple isoforms. *PLoS One.* 2012; 7(10): e48745. <https://doi.org/10.1371/journal.pone.0048745> PMID: 23119097
17. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.* 2011; 12(1):R6. <https://doi.org/10.1186/gb-2011-12-1-r6> PMID: 21247443
18. Schulte I, Batty EM, Pole JC, Blood KA, Mo S, Cooke SL, et al. Structural analysis of the genome of breast cancer cell line ZR-75-30 identifies twelve expressed fusion genes. *BMC Genomics.* 2012; 13:719. <https://doi.org/10.1186/1471-2164-13-719> PMID: 23260012
19. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* 2009; 462(7276):1005–10. <https://doi.org/10.1038/nature08645> PMID: 20033038
20. Hampton OA, Den Hollander P, Miller CA, Delgado DA, Li J, Coarfa C, et al. A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res.* 2009; 19(2):167–77. <https://doi.org/10.1101/gr.080259.108> PMID: 19056696
21. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.* 2013; 14(2):R12. <https://doi.org/10.1186/gb-2013-14-2-r12> PMID: 23409703
22. Gudmundsdottir ET, Barkardottir RB, Arason A, Gunnarsson H, Amundadottir LT, Agnarsson BA, et al. The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. *BMC Cancer.* 2012; 12:621. <https://doi.org/10.1186/1471-2407-12-621> PMID: 23270421
23. Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Holm K, Hegardt C, et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res.* 2010; 12(3):R42. <https://doi.org/10.1186/bcr2596> PMID: 20576095
24. Reynisdottir I, Arason A, Einarsdottir BO, Gunnarsson H, Staaf J, Vallon-Christersson J, et al. High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. *Cancer Med.* 2013; 2(4):437–46. <https://doi.org/10.1002/cam4.88> PMID: 24156016
25. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell.* 2015; 163(2):506–19. <https://doi.org/10.1016/j.cell.2015.09.033> PMID: 26451490
26. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012; 486(7403):346–52. <https://doi.org/10.1038/nature10983> PMID: 22522925
27. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016; 7:11479. <https://doi.org/10.1038/ncomms11479> PMID: 27161491
28. Rueda OM, Sammut SJ, Seoane JA, Chin SF, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature.* 2019; 567(7748):399–404. <https://doi.org/10.1038/s41586-019-1007-8> PMID: 30867590
29. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2(5):401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
30. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6(269):pl1. <https://doi.org/10.1126/scisignal.2004088> PMID: 23550210
31. Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Gruvberger-Saal SK, Saal LH, et al. The retinoblastoma gene undergoes rearrangements in BRCA1-deficient basal-like breast cancer. *Cancer Res.* 2012; 72(16):4028–36. <https://doi.org/10.1158/0008-5472.CAN-12-0097> PMID: 22706203
32. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490(7418):61–70. <https://doi.org/10.1038/nature11412> PMID: 23000897

33. R. The R Project for Statistical Computing [Available from: <https://www.r-project.org/>].
34. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis Part III: multivariate data analysis—choosing a model and assessing its adequacy and fit. *Br J Cancer*. 2003; 89(4):605–11. <https://doi.org/10.1038/sj.bjc.6601120> PMID: 12915864
35. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*. 2009; 462(7276):1005–10. <https://doi.org/10.1038/nature08645> PMID: 20033038
36. Inaki K, Hillmer AM, Ukil L, Yao F, Woo XY, Vardy LA, et al. Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome Res*. 2011; 21(5):676–87. <https://doi.org/10.1101/gr.113225.110> PMID: 21467264
37. Pérez-Tenorio G, Karlsson E, Waltersson MA, Olsson B, Holmlund B, Nordenskjöld B, et al. Clinical potential of the mTOR targets S6K1 and S6K2 in breast cancer. *Breast Cancer Res Treat*. 2011; 128(3):713–23. <https://doi.org/10.1007/s10549-010-1058-x> PMID: 20953835
38. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*. 2006; 7:96. <https://doi.org/10.1186/1471-2164-7-96> PMID: 16643655
39. Andersen CL, Monni O, Wagner U, Kononen J, Barlund M, Bucher C, et al. High-throughput copy number analysis of 17q23 in 3520 tissue specimens by fluorescence in situ hybridization to tissue microarrays. *Am J Pathol*. 2002; 161(1):73–9. [https://doi.org/10.1016/S0002-9440\(10\)64158-2](https://doi.org/10.1016/S0002-9440(10)64158-2) PMID: 12107091
40. Haverty PM, Fridlyand J, Li L, Getz G, Beroukhir R, Lohr S, et al. High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer*. 2008; 47(6):530–42. <https://doi.org/10.1002/gcc.20558> PMID: 18335499
41. Natrajan R, Lambros MB, Rodríguez-Pinilla SM, Moreno-Bueno G, Tan DS, Marchió C, et al. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res*. 2009; 15(8):2711–22. <https://doi.org/10.1158/1078-0432.CCR-08-1878> PMID: 19318498
42. Ribas J, Ni X, Castanares M, Liu MM, Esopi D, Yegnasubramanian S, et al. A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res*. 2012; 40(14):6821–33. <https://doi.org/10.1093/nar/gks308> PMID: 22505577
43. Fujita S, Ito T, Mizutani T, Minoguchi S, Yamamichi N, Sakurai K, et al. miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *J Mol Biol*. 2008; 378(3):492–504. <https://doi.org/10.1016/j.jmb.2008.03.015> PMID: 18384814
44. Staaf J, Jonsson G, Ringner M, Vallon-Christersson J, Grabau D, Arason A, et al. High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res*. 2010; 12(3):R25. <https://doi.org/10.1186/bcr2568> PMID: 20459607
45. Molejon MI, Ropolo A, Vaccaro MI. VMP1 is a new player in the regulation of the autophagy-specific phosphatidylinositol 3-kinase complex activation. *Autophagy*. 2013; 9(6):933–5. <https://doi.org/10.4161/autophagy.24390> PMID: 23558782
46. Galluzzi L, Pietrocola F, Bravo-San Pedro JM, Amaravadi RK, Baehrecke EH, Cecconi F, et al. Autophagy in malignant transformation and cancer progression. *EMBO J*. 2015; 34(7):856–80. <https://doi.org/10.15252/embj.201490784> PMID: 25712477
47. Lo Ré AE, Fernández-Barrena MG, Almada LL, Mills LD, Elswa SF, Lund G, et al. Novel AKT1-GLI3-VMP1 pathway mediates KRAS oncogene-induced autophagy in cancer cells. *J Biol Chem*. 2012; 287(30):25325–34. <https://doi.org/10.1074/jbc.M112.370809> PMID: 22535956
48. Rodríguez ME, Catrinacio C, Ropolo A, Rivarola VA, Vaccaro MI. A novel HIF-1 α /VMP1-autophagic pathway induces resistance to photodynamic therapy in colon cancer cells. *Photochem Photobiol Sci*. 2017; 16(11):1631–42. <https://doi.org/10.1039/c7pp00161d> PMID: 28936522
49. Zheng L, Chen L, Zhang X, Zhan J, Chen J. TMEM49-related apoptosis and metastasis in ovarian cancer and regulated cell death. *Mol Cell Biochem*. 2016; 416(1–2):1–9. <https://doi.org/10.1007/s11010-016-2684-3> PMID: 27023910
50. Tabara LC, Vicente JJ, Biazik J, Eskelinen EL, Vincent O, Escalante R. Vacuole membrane protein 1 marks endoplasmic reticulum subdomains enriched in phospholipid synthesizing enzymes and is required for phosphoinositide distribution. *Traffic*. 2018; 19(8):624–38. <https://doi.org/10.1111/tra.12581> PMID: 29761602
51. Sauermaun M, Sahin O, Sultmann H, Hahne F, Blaszkiewicz S, Majety M, et al. Reduced expression of vacuole membrane protein 1 affects the invasion capacity of tumor cells. *Oncogene*. 2008; 27(9):1320–6. <https://doi.org/10.1038/sj.onc.1210743> PMID: 17724469
52. Zhao YG, Chen Y, Miao G, Zhao H, Qu W, Li D, et al. The ER-Localized Transmembrane Protein EPG-3/VMP1 Regulates SERCA Activity to Control ER-Isolation Membrane Contacts for Autophagosome

- Formation. *Mol Cell*. 2017; 67(6):974–89 e6. <https://doi.org/10.1016/j.molcel.2017.08.005> PMID: 28890335
53. Tabara LC, Vincent O, Escalante R. Evidence for an evolutionary relationship between Vmp1 and bacterial DedA proteins. *Int J Dev Biol*. 2019.
 54. Guo L, Yang LY, Fan C, Chen GD, Wu F. Novel roles of Vmp1: inhibition metastasis and proliferation of hepatocellular carcinoma. *Cancer Sci*. 2012; 103(12):2110–9. <https://doi.org/10.1111/cas.12025> PMID: 22971212
 55. Guo XZ, Ye XL, Xiao WZ, Wei XN, You QH, Che XH, et al. Downregulation of VMP1 confers aggressive properties to colorectal cancer. *Oncol Rep*. 2015; 34(5):2557–66. <https://doi.org/10.3892/or.2015.4240> PMID: 26328607
 56. Singh SS, Vats S, Chia AY, Tan TZ, Deng S, Ong MS, et al. Dual role of autophagy in hallmarks of cancer. *Oncogene*. 2017.
 57. Molejon MI, Ropolo A, Re AL, Boggio V, Vaccaro MI. The VMP1-Beclin 1 interaction regulates autophagy induction. *Sci Rep*. 2013; 3:1055. <https://doi.org/10.1038/srep01055> PMID: 23316280
 58. Vega-Rubín-de-Celis S, Zou Z, Fernández Á, Ci B, Kim M, Xiao G, et al. Increased autophagy blocks HER2-mediated breast tumorigenesis. *Proc Natl Acad Sci U S A*. 2018; 115(16):4176–81. <https://doi.org/10.1073/pnas.1717800115> PMID: 29610308
 59. Loncle C, Molejon MI, Lac S, Tellechea JI, Lomberk G, Gramatica L, et al. The pancreatitis-associated protein VMP1, a key regulator of inducible autophagy, promotes Kras(G12D)-mediated pancreatic cancer initiation. *Cell Death Dis*. 2016; 7:e2295. <https://doi.org/10.1038/cddis.2016.202> PMID: 27415425
 60. Jarman EJ, Ward C, Turnbull AK, Martínez-Perez C, Meehan J, Xintaropoulou C, et al. HER2 regulates HIF-2 α and drives an increased hypoxic response in breast cancer. *Breast Cancer Res*. 2019; 21(1):10. <https://doi.org/10.1186/s13058-019-1097-0> PMID: 30670058
 61. Jacot W, Fiche M, Zaman K, Wolfer A, Lamy PJ. The HER2 amplicon in breast cancer: Topoisomerase IIA and beyond. *Biochim Biophys Acta*. 2013; 1836(1):146–57. <https://doi.org/10.1016/j.bbcan.2013.04.004> PMID: 23628726
 62. Sahlberg KK, Hongisto V, Edgren H, Makela R, Hellstrom K, Due EU, et al. The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells. *Mol Oncol*. 2013; 7(3):392–401. <https://doi.org/10.1016/j.molonc.2012.10.012> PMID: 23253899
 63. Ren L, Liu J, Gou K, Xing C. Copy number variation and high expression of DNA topoisomerase II alpha predict worse prognosis of cancer: a meta-analysis. *J Cancer*. 2018; 9(12):2082–92. <https://doi.org/10.7150/jca.23681> PMID: 29937926
 64. Jasra S, Anampa J. Anthracycline Use for Early Stage Breast Cancer in the Modern Era: a Review. *Curr Treat Options Oncol*. 2018; 19(6):30. <https://doi.org/10.1007/s11864-018-0547-8> PMID: 29752560
 65. Liu Y, Xu J, Choi HH, Han C, Fang Y, Li Y, et al. Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer. *Nat Commun*. 2018; 9(1):4718. <https://doi.org/10.1038/s41467-018-07264-0> PMID: 30413718
 66. Gilbert M, Vaccaro MI, Fernandez-Zapico ME, Calvo EL, Turrini O, Secq V, et al. Novel role of VMP1 as modifier of the pancreatic tumor cell response to chemotherapeutic drugs. *J Cell Physiol*. 2013; 228(9):1834–43. <https://doi.org/10.1002/jcp.24343> PMID: 23460482
 67. Zhou Y, Rucker EB, Zhou BP. Autophagy regulation in the development and treatment of breast cancer. *Acta Biochim Biophys Sin (Shanghai)*. 2016; 48(1):60–74.
 68. Vazquez-Martin A, Oliveras-Ferraro C, Menendez JA. Autophagy facilitates the development of breast cancer resistance to the anti-HER2 monoclonal antibody trastuzumab. *PLoS One*. 2009; 4(7):e6251. <https://doi.org/10.1371/journal.pone.0006251> PMID: 19606230
 69. Chen S, Zhu X, Qiao H, Ye M, Lai X, Yu S, et al. Protective autophagy promotes the resistance of HER2-positive breast cancer cells to lapatinib. *Tumour Biol*. 2016; 37(2):2321–31. <https://doi.org/10.1007/s13277-015-3800-9> PMID: 26369543
 70. Samaddar JS, Gaddy VT, Duplantier J, Thandavan SP, Shah M, Smith MJ, et al. A role for macroautophagy in protection against 4-hydroxytamoxifen-induced cell death and the development of antiestrogen resistance. *Mol Cancer Ther*. 2008; 7(9):2977–87. <https://doi.org/10.1158/1535-7163.MCT-08-0447> PMID: 18790778
 71. Qadir MA, Kwok B, Dragowska WH, To KH, Le D, Bally MB, et al. Macroautophagy inhibition sensitizes tamoxifen-resistant breast cancer cells and enhances mitochondrial depolarization. *Breast Cancer Res Treat*. 2008; 112(3):389–403. <https://doi.org/10.1007/s10549-007-9873-4> PMID: 18172760

Paper II

Hsa-miR-21-3p associates with breast cancer patient survival and targets genes in tumor suppressive pathways

Arsalan Amirfallah ^{1,2}, Hildur Knutsdottir ³, Adalgeir Arason ^{4,5}, Bylgja Hilmarsdottir ^{6,7}, Oskar T. Johannsson ⁸, Bjarni A. Agnarsson ^{9,10}, Rosa B. Barkardottir ^{11,12} and Inga Reynisdottir ^{13,14, *}

¹ Cell Biology Unit, Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; arsalan@landspitali.is

² BMC (Biomedical Center), Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; ara54@hi.is

³ Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218; hildur@jhu.edu

⁴ Molecular Pathology Unit, Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; adalgeir@landspitali.is

⁵ BMC (Biomedical Center), Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; adalgeir@landspitali.is

⁶ Molecular Pathology Unit, Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; bylgjahi@landspitali.is

⁷ BMC (Biomedical Center), Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; bylgjahi@landspitali.is

⁸ Department of Oncology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; oskarjoh@landspitali.is

⁹ Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; bjarniaa@landspitali.is

¹⁰ Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; bjarniaa@landspitali.is

¹¹ Molecular Pathology Unit, Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; rosa@landspitali.is

¹² BMC (Biomedical Center), Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; rosa@landspitali.is

¹³ Cell Biology Unit, Department of Pathology, Landspítali – The National University Hospital of Iceland, 101 Reykjavik, Iceland; ingar@landspitali.is

¹⁴ BMC (Biomedical Centre), Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; ingareynis@hi.is

* Correspondence: ingar@landspitali.is

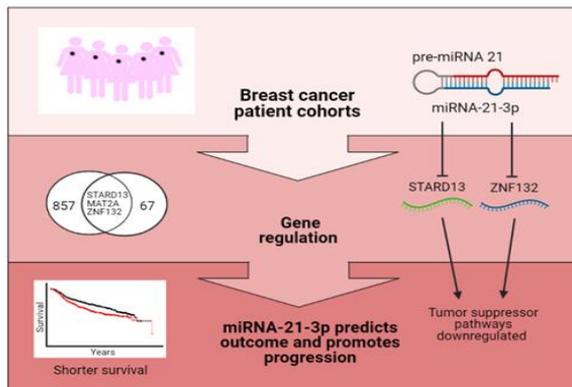
Simple Summary: Breast cancer is the most common cancer among women world-wide. Although the five-year survival rate is high, not all respond to therapy and some relapse after treatment. In breast cancer patients, high levels of the oncomiR, microRNA-21 (MIR21), indicate shorter survival. MIR21 produces two mature products, miR-21-5p, which has been studied in detail, and miR-21-3p, which has received less attention. Our previous study showed miR-21-3p was highly expressed in breast tumors, suggesting miR-21-3p overexpression contributes to breast cancer progression. Here, we investigated whether miR-21-3p expression predicts survival and identified cellular pathways modulated by miR-21-3p expression. Our results show miR-21-3p controls pathways that support cancer progression and suggest both new markers for predicting patient prognosis and new potential drug targets.

Abstract: MicroRNA-21 (MIR21) is a well-studied oncomiR in breast cancer with most of its effects attributed to hsa-miR-21-5p. Our aim here was twofold: analyze whether expression of the less studied hsa-miR-21-3p (miR-21-3p) is similarly prognostic for breast cancer and use bioinformatics tools to infer its function. MiR-21-3p association with survival, clinical and pathological characteristics was analyzed in a breast cancer cohort (cohort-1) and validated in separate cohorts (cohort-2, TCGA, and METABRIC). Correlation analysis between miR-21-3p and mRNA expression identified potential target genes and functional pathways. In cohort-1, high miR-21-3p levels associated with shorter survival and lymph node positivity. In the largest validation cohort, METABRIC (n=1174), high miR-21-3p levels associated with large tumors, a high grade, lymph node and HER2 positivity, and shorter breast-cancer-specific survival (HR=1.38, CI 1.13-1.68). This association remained significant after adjusting for confounding factors. The genes with expression levels that correlated with miR-21-3p were enriched in particular pathways, including the epithelial-to-mesenchymal transition and proliferation. The most significantly down-regulated targets were MAT2A and the tumor suppressive genes STARD13 and ZNF132. Our data suggest miR-21-3p overexpression in breast tumors is a marker of breast cancer

progression. Furthermore, miR-21-3p affects genes in pathways that are active in tumor cells and might drive breast cancer by down-regulating tumor suppressor genes.

Keywords: MIR21, miR-21-3p, breast cancer, survival, prognosis, targets, pathways.

1. Introduction



MiRs are small non-coding RNAs that regulate gene expression by binding to 3'UTRs of target mRNAs to cause them to be unstable and/or degrade. MiRs are transcribed as long primary miRs (pri-miRs) and subsequently processed to much shorter pre-miRs that can give rise to two mature molecules miR-5p and/or miR-3p [1]. MiRs can be located both intra- and intergenically and can be transcribed independently from their own promoter or the promoter of the gene in which they reside [1].

MIR21 overlaps with the vacuole membrane protein 1 (VMP1), sharing about 1 kb of sequence [2]. The two genes have separate promoters, with pri-miR-21 transcribed from its own promoter, miPPR-21 [3,4]. That promoter is a site of active transcription, as it contains binding sites for transcription factors such as STAT3, AP-1, C/EBP and p53 among others [5] (reviewed in [6]). Both VMP1 and MIR21 have their own polyadenylation sites, located upstream of the miR-21 hairpin for VMP1 but downstream of the hairpin for MIR21 [2,7]. This suggests that the two genes are transcribed independent of each other.

MIR21 is well studied in cancer and is reported as a potential diagnostic, prognostic, and predictive biomarker in many cancer types, including breast cancer (reviewed in [6]). A meta-analysis of breast cancer demonstrated that elevated levels of miR-21 predict poor prognosis for breast cancer patients when miR-21 was measured in breast tumors and in tumor cells circulating in serum [8]. Most clinical studies that analyzed miR-21 (see studies within this meta-analysis [9]) used probes that measured miR-21 (miR-21-5p) but not miR-21* (miR-21-3p). Overexpression of miR-21 was reported to enhance cellular proliferation and induce invasion and metastasis. This was achieved through interaction with its target genes, many of which are known tumor suppressor genes, e.g., division cycle 25A (CDC25A), programmed cell death 4 (PDCD4), tropomyosin 1 (TPM1) and phosphatase and tensin homolog (PTEN) [10-13]. Only a small number of studies focused on the role of miR-21-3p in breast cancer survival, with conflicting results.

MiR-21-3p levels were reportedly higher in breast tumors than in normal breast tissue [14,15]. In a clinical study performed using triple negative breast tumors from The Cancer Genome Atlas (TCGA), low miR-21-3p expression associated with shorter overall survival and, when combined in a panel with two other miRs, was suggested to be a prognostic marker predicting shorter survival [14]. In another study using breast tissue and serum samples from TCGA, high miR-21-3p expression was suggested to be a non-invasive prospective marker for detection of early-stage breast cancer (when it was used in a panel with two other miRs) [16]. Another study found that, when analyzing miR-21-3p paired with another miR, the pair could distinguish between breast tumor tissue and benign lesions [17]. Although miR-21-3p studies in breast cancer are few, emerging evidence suggests it plays a role in the disease. Thus, the focus of this study is miR-21-3p and its potential role as a biomarker in breast cancer. Here, we examined whether miR-21-3p expression was associated with tumor malignancy, shorter survival, and a poor prognosis for breast cancer patients. In addition, our bioinformatic analyses sought pathways and targets modulated by miR-21-3p.

Our results show that elevated miR-21-3p expression associated with clinical and pathological characteristics that indicate disease severity and shorter breast-cancer-specific survival. Bioinformatic analyses revealed that genes positively correlating with miR-21-3p expression were enriched in pathways that induce proliferation and the epithelial-to-mesenchymal transition, whereas genes that negatively correlated included candidates that when underexpressed were implicated in tumor progression.

2. Materials and Methods

2.1 Cohorts and clinical data

Breast cancer patients in cohort-1 (n = 158; diagnosed 1987 to 2003) and cohort-2 (n = 291; diagnosed 2003 – 2007), and the collection of relevant patient and tumor data have been reported [18,19]. Primary, fresh frozen tumors and normal breast tissue (non-neoplastic breast tissue dissected as far away from the tumor as possible) were obtained from the Department of Pathology. Informed consent was obtained from all subjects involved. The study was approved by the National Bioethics Committee of Iceland (VSN-15-138). Data for the breast cancer patients from TCGA cohort (Firehose Legacy, n = 1108) [20], diagnosed 1988 – 2013, and the METABRIC cohort (n = 2509) [21-23], diagnosed 1980 – 2005, were collected through the cBioPortal [24,25].

2.2 DNA and RNA isolation

DNA and total RNA were extracted from 35 fresh frozen non-neoplastic breast tissue samples from cohort-2 using the Allprep kit DNA/RNA/miRNA (Qiagen no. 80224). Since normal breast tissue is enriched in fatty tissue and stromal cells, as compared to the tumor, a section of normal breast tissue was stained with eosin and hematoxylin, before extraction, to ensure the presence of normal epithelial breast cells in the specimen. DNA and total RNA was extracted from cohort-2 by the same method used for normal breast tissue [18], but total RNA from cohort-1 was extracted with Trizol, as described [19].

2.3 MiRNA and mRNA data

MiR-21 expression was quantified in breast tumors from cohorts-1 and 2. A miRCURY LNA RT Kit (Qiagen) was used to generate cDNA from breast tumors and normal breast tissue from cohorts-1 and 2 by following the manufacturer's protocol. Five ng/μl of RNA were used. Quantitative-PCR (qPCR) was performed using miRCURY LNA SYBR Green PCR Kit (Qiagen) and miRCURY primer sets for hsa-miR-21-3p (YP00204302) and hsa-miR-16-5p (YP00205702), with the latter as a reference gene. The assay quantifies the reference isomiRs. Reactions were performed in triplicate using 40 cycles according to the manufacturer's protocol. Values for miR-21-3p and mRNAs from TCGA were retrieved through the cBioPortal [24,25]. Two libraries, Illumina Genome Analyzer (238 patients) and Illumina HiSeq 2000 (717 patients), were used to generate the RNA-Seq data that are the basis for the miR-21-3p isomiR extraction [20]. Nine isomiRs representing miR-21-

3p were used, based on their differential expression, higher in tumors than normal (Figure S1 and Table S1). MiR-21-3p expression values from METABRIC were not available through the cBioPortal and so were retrieved from the European Genome-phenome Archive (EGA: <https://ega-archive.org/>) from study EGAS00000000122, dataset accession number EGAD00010000438. The probe used to detect miR-21-3p on the Agilent microarray is 15 nucleotides and binds the reference isomiR sequence, and possibly other isomiRs of miR-21-3p [26]. MiR-21-5p expression values were also from EGA. For the bioinformatic analyses, normalized mRNA values for the METABRIC cohort were retrieved from the same study (dataset accession EGAD00010000434).

2.4 Statistical analyses

Patients that lacked data for mRNA, miRNA, or survival were excluded from the analyses. The number of patients in each cohort was 139, 281, 946, and 1174 for cohort-1, cohort-2, TCGA, and METABRIC, respectively. The miR-21-3p values from cohorts-1 and 2 and TCGA were transformed with log₂ to normalize the data. The miR-21-3p and miR-21-5p values from METABRIC had been normalized [27]. All miR values were centered at 0. VMP1, TUBD1, RPS6KB1, and PPM1D mRNAs from METABRIC that were retrieved from the cBioPortal had been normalized (Z scores). The statistical program R version 3.5.3 was used for the analyses [28]. Correlation between miRNA and mRNA expression was calculated by Pearson's product moment correlation using normalized values. For the bioinformatic analyses, Stouffer's method for meta-analysis was used to combine results from TCGA and METABRIC, $Z = \frac{\sum_{i=1}^k Z_i}{\sqrt{k}}$, where Z are the Z-scores and k=2 representing TCGA and METABRIC. Genes were considered significantly correlated using the 5% FDR (false discovery rate) threshold in each cohort and the more stringent 5% FWER (familywise error rate) threshold in the meta-analysis. The association of miR-21-3p with clinical and pathological characteristics was performed with Student's t-test or ANOVA. Expression levels in breast tumors and normal breast tissue were compared with a paired t-test. Kaplan-Meier and log-rank tests were calculated to estimate survival using the survival and survminer packages in R. Tumors were classified into high- and low-expressing tumors, based on the median miR-21-3p expression values in each cohort. Cox regression analyses calculated hazard ratios (HR) and the effect of tumor characteristics with an independent effect on survival. Characteristics with numerical values were analyzed as both categorical and continuous variables. P-values below 0.05 were considered significant.

3. Results

3.1 *MiR-21-3p* associates with metastasis and shorter disease-free survival

The relationship between miR-21-3p and clinical and pathological characteristics was first explored in cohort-1, a breast cancer cohort that contains 139 patients. In breast tumors from cohort-1, RT-qPCR measured miR-21-3p expression and the resultant values were used to test whether miR-21-3p expression levels associated with any clinical and pathological characteristics. This analysis showed miR-21-3p expression was significantly higher in breast tumors from patients with metastasis than in patients without ($p = 2.1 \cdot 10^{-2}$). No statistically significant correlation with other clinical and pathological parameters was detected (**Table S2**) but, in cohort-1, patients expressing high (above median) levels of miR-21-3p had a significantly shorter disease-free survival (DFS; log rank $p = 7.1 \cdot 10^{-3}$; **Figure 1**).

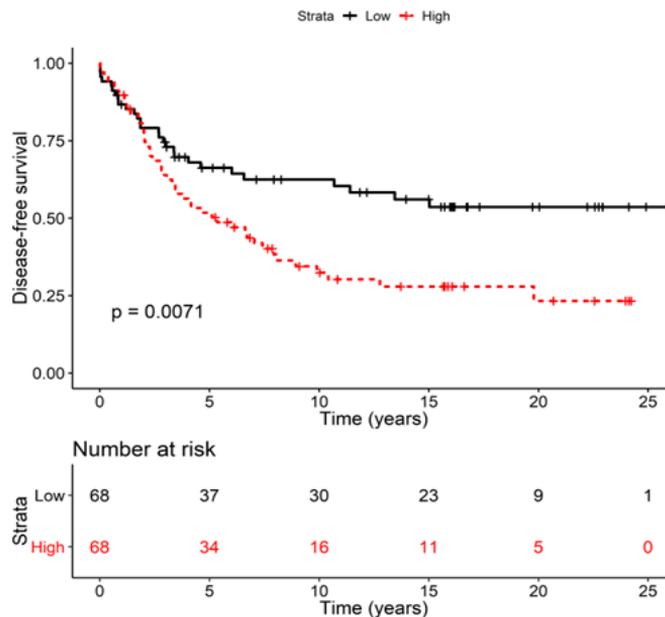


Figure 1. High miR-21-3p levels associated with shorter DFS. Disease-free survival (DFS) was examined in cohort 1, an exploration cohort. Patients were divided into two groups based on median expression of miR-21-3p; high reflects above the median expression (red) and low reflects below median expression (black). The log rank p-value was $7.1 \cdot 10^{-3}$. The number of patients at risk at the indicated time point is shown in a table below the graph. The

DFS HR was 1.89 (CI 1.18 – 3.04) and after adjusting for HER2 the HR was 1.72 (CI 1.08 – 2.78).

The effect of miR-21-3p association on DFS was assessed with Cox regression analysis. The HR of miR-21-3p was 1.89 (95% CI 1.18-3.04). Among breast tumors with amplified ERBB2 (the gene that expresses the HER2 receptor), 30% were also amplified for the MIR21 genomic region [29]. Examining the effect of HER2 expression on miR-21-3p associated survival showed that HR was 1.72 (95% CI 1.08-2.78) after adjusting for HER2 expression. This shows that HER2 expression attenuates the association of miR-21-3p with DFS. Nevertheless, the effect remained significant suggesting high miR-21-3p expression in breast tumors affects the recurrence rate of breast cancer.

3.2 MiR-21-3p associates with tumor characteristics that indicate worse prognosis

To follow up results in cohort-1, association with clinicopathological characteristics and survival was analyzed in cohort-2 (n = 281), TCGA (n = 946) and METABRIC (n = 1174). In METABRIC, miR-21-3p was highly expressed in HER2-positive tumors ($p = 2.63 \cdot 10^{-9}$), large tumors (> 20 mm, $p = 2.0 \cdot 10^{-2}$), tumors of histologic grade 3 ($p = 3.68 \cdot 10^{-14}$), lymph node-positive tumors ($p = 1.0 \cdot 10^{-3}$), HER2 tumors according to PAM50 classification ($p < 2 \cdot 10^{-16}$), and HER2+ tumors according to the 3-Gene classifier subtype ($p < 2 \cdot 10^{-16}$; **Table S3**). Furthermore, miR-21-3p was more highly expressed in ductal than lobular tumors ($1.69 \cdot 10^{-7}$) and in tumors from patients, who had developed metastasis ($p = 2.0 \cdot 10^{-2}$). Compared to METABRIC, both cohort-2 and TCGA represent fewer tumors and fewer available clinical and pathological parameters. Even so, the data confirmed the association detected in METABRIC. MiR-21-3p values were again significantly higher in HER2-positive tumors in cohort-2 ($p = 3.0 \cdot 10^{-3}$) and in histologic grade 3 tumors ($p = 2.6 \cdot 10^{-2}$; **Table S4**). Also, in cohort-2, miR-21-3p was higher in large tumors (> 20 mm, $p = 7.8 \cdot 10^{-2}$) and in tumors from patients with metastasis ($p = 6.8 \cdot 10^{-2}$). In TCGA, miR-21-3p was more highly expressed in HER2-positive tumors ($p = 5.82 \cdot 10^{-4}$) and the HER2-enriched molecular subtype ($p = < 2 \cdot 10^{-16}$; **Table S5**). Taken together, these data suggest high miR-21-3p levels associated with clinical and pathological characteristics of worse patient prospects.

3.3 High miR-21-3p expression associates with shorter breast cancer specific survival

Since malignant tumor characteristics likely predict survival outcomes, breast cancer-specific survival (BCSS), as a function of miR-21-3p expression, was calculated for cohort-2, TCGA, and METABRIC. In METABRIC, patients overexpressing miR-21-3p (above median) had significantly shorter BCSS than those expressing levels below the median (log rank $p = 1.6 \cdot 10^{-3}$, **Figure 3**). In cohort-2, BCSS was borderline significant (log rank $p = 6.0 \cdot 10^{-2}$; **Figure S2a**), while in TCGA there was no association (**Figure S2b**).

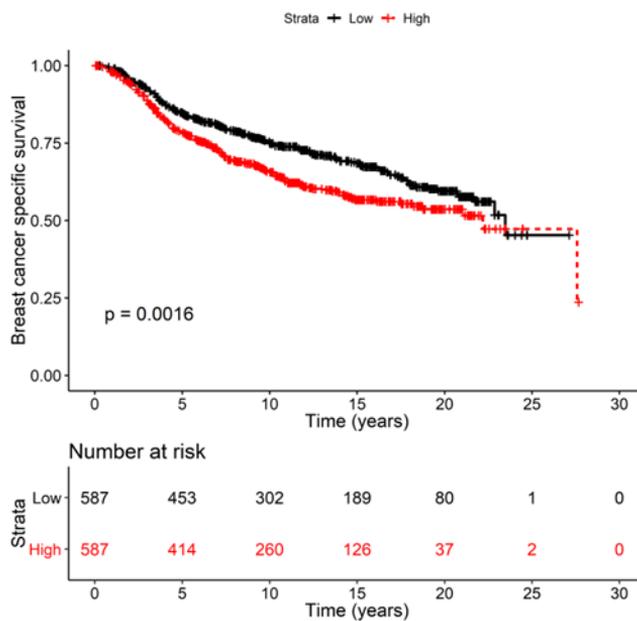


Figure 2. High miR-21-3p expression associated with shorter BCSS in METABRIC. Breast cancer-specific survival (BCSS) was examined in METABRIC, the largest validation cohort. Patients were divided into two groups according to expression of miR-21-3p: above median (red) and below median (black). The log rank p -value was $1.6 \cdot 10^{-3}$. The number of patients at risk at each time point is shown in a table below the graph. The HR was 1.39 (CI 1.15 – 1.70).

Submitted manuscript

Differences in the composition, time of diagnosis and treatment of the cohorts may contribute to the inconsistent results. MiR-21-3p expression was high in grade 3 tumors, large tumors, and tumors with lymph node positivity, all of which predict a poorer outcome. A Cox regression analysis tested whether high miR21-3p expression correlated with shorter survival in METABRIC because of characteristics other than miR-21-3p. The HR of miR-21-3p on BCSS was 1.38 (95% CI 1.13 – 1.68; **Table 1**). Lymph node positivity and high tumor stage attenuated the effect of miR-21-3p slightly, but the largest confounders were HER2 and histologic grade. If HER2 positivity was considered, then the HR of miR-21-3p was reduced to 1.28 (95% CI 1.05 – 1.57), and with histologic grade the HR of miR-21-3p decreased to 1.23 (95% CI 1.01 – 1.51). Nevertheless, even after adjusting for confounding variables, the effect of miR-21-3p on survival remained, suggesting high levels of miR-21-3p contribute to a worse prognosis.

Table 1. Adjustment of BCSS in METABRIC (n = 1174) for confounding variables

	HR	CI	p-value
miR-21-3p	1.377	1.127 – 1.681	2.0·10 ⁻³
+HER2	1.284	1.048 – 1.574	1.6·10 ⁻²
+ER	1.336	1.093 – 1.663	5.0·10 ⁻³
+PR	1.336	1.094 – 1.633	5.0·10 ⁻³
+age ¹	1.377	1.128 – 1.682	2.0·10 ⁻³
+tumor size	1.416	1.158 – 1.731	7.0·10 ⁻⁴
+nodes	1.319	1.080 – 1.611	7.0·10 ⁻³
+ grade	1.231	1.006 – 1.508	4.4·10 ⁻²
+VMP1 ¹	1.470	1.179 – 1.833	6.0·10 ⁻⁴
+RPS6KB1 ¹	1.385	1.129 – 1.700	2.0·10 ⁻³
+PPM1D ¹	1.403	1.146 – 1.718	1.0·10 ⁻³
+miR-21-5p ¹	1.377	1.126 – 1.684	2.0·10 ⁻³

3.4 MiR-21-3p is more highly expressed in breast tumors than normal breast tissue

One indication that miR-21-3p plays a role in tumor progression might be a difference in expression level when comparing normal and tumor tissue. To test this, RT-qPCR measured miR-21-3p levels, comparing breast tumor to normal breast tissue, in samples from 35 patients from cohort-2. MiR-21-3p

expression was significantly higher in tumors (paired t-test, $p = 4.50 \cdot 10^{-13}$, **Figure S3a**), a result confirmed by comparing miR-21-3p expression levels in 172 breast tumors from The Cancer Genome Atlas (TCGA) with matched normal breast tissue ($p = 1.10 \cdot 10^{-15}$, **Figure S3b**).

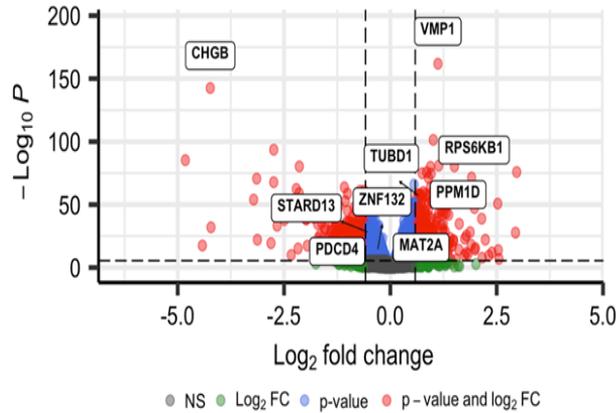
3.5 Genes co-amplified with miR-21-3p did not attenuate its effect on survival

MIR21 is located at 17q23.1 (GRCh38), a region frequently amplified in breast tumors. Depending on the tumor's histological origin, this region can be amplified in up to 22% of primary breast tumors [30]. Genes in amplified regions are sometimes overexpressed, which can support tumor development. High expression from genes neighboring MIR21 has been reported in breast tumors [31]; and some genes, including RPS6KB1, PPM1D and VMP1, have been implicated in tumor development [18,32,33]. Therefore, we performed correlation analyses with data from METABRIC to examine whether miR-21-3p was highly expressed in concurrence with RPS6KB1, PPM1D and VMP1. Expression of miR-21-3p significantly correlated with RPS6KB1 ($r = 0.38$, $p < 2.2 \cdot 10^{-16}$), PPM1D ($r = 0.31$, $p < 2.2 \cdot 10^{-16}$), and VMP1 ($r = 0.57$, $p < 2.2 \cdot 10^{-16}$) (**Figure S4a-c**). Even so, high expression from these genes did not attenuate miR-21-3p's effect on survival (**Table 1**). The correlation with miR-21-5p was analyzed as well. In the METABRIC cohort, there was a significant correlation between miR-21-3p and miR-21-5p expression ($r = 0.1$, $p = 4.89 \cdot 10^{-4}$; **Figure S4d**). Nevertheless, the elevation of miR-21-5p expression level was incremental and the effect of miR-21-3p on survival was not confounded by miR-21-5p (**Table 1**).

3.6 MiR-21-3p down-regulates potential tumor suppressor genes

MiRs bind their target mRNAs, destabilizing them which results in their degradation. To identify targets of miR-21-3p, we conducted a correlation analysis in the METABRIC and TCGA cohorts and combined the results in a meta-analysis using Stouffer's method. A volcano plot of the results (~12,400 genes), showing fold change ($\log_2(\text{FC})$) and the significance values (negative $\log_{10}P$) was plotted (**Figure 3a**). In total, we identified 853 down-regulated genes and 1,822 up-regulated genes at 5% FWER for the meta-analysis and further requiring that each gene also be significantly correlated with miR-21-3p in each cohort (using the less stringent 5% FDR for significance cut-off).

a)



b)

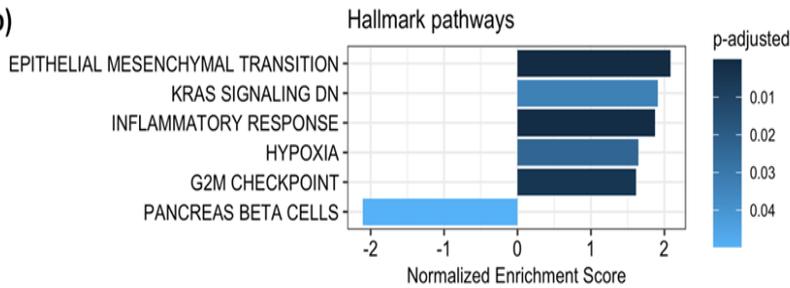


Figure 3. Meta-analysis of mRNA correlation with miR-21-3p. (a) Volcano plot from TCGA and METABRIC meta-analysis, showing genes in red that had a significant p-value (FWER threshold) and at least 1.5-fold change in expression. Select genes relevant to this research are highlighted. (b) Gene set enrichment analysis (GSEA) of the Hallmark pathways showed that genes that positively correlate with miR-21-3p expression fall within pathways of EMT, proliferation and inflammation while genes that negatively correlate with miR-21-3p are in the pancreas beta cells pathway.

Not surprisingly, the upregulated gene that correlated most significantly with miR-21-3p was VMP1 (up- and down-regulated genes are listed in **Table S6**). Gene set enrichment analysis (GSEA) found miR-21-3p expression positively correlated most with its neighboring genes within the amplicon at 17q23.1, namely Farmer’s cluster 5 (p-adjusted = $2 \cdot 10^{-23}$; [34]) and genes in amplicon 17q21-25 (p-adjusted = $1 \cdot 10^{-21}$; [35]; **Table S7**). GSEA of the Hallmark pathways identified pathways that support proliferation, the epithelial-to-mesenchymal transition (EMT), and responses to inflammation (**Figure 3b**). Since miR-21-3p expression correlated with expression of its neighboring genes, these data cannot distinguish whether the downstream effects are due to miR-21-3p or neighboring genes. Most likely, elevated expression from all of them is a contributing factor.

Gene ontology (GO) gene sets that correlated significantly with genes that inversely correlate with miR-21-3p include metabolic processes, transmembrane transport, and cilium organization. Cilium organization is a key signaling hub, for example in Wnt and MAPK signaling, and plays a role in cancer [36] (**Table S8**). The most down-regulated gene from the meta-analysis was chromogranin-B (CHGB), which is associated with malignancy and metastasis in pancreatic neuroendocrine tumors (PNETs) [37]. Lower expression of CHGB was reported in invasive ductal carcinoma of the breast as compared to non-invasive ductal carcinoma [38]; and breast cancer patients with CHGB negative tumors have poorer prognosis than those with CHGB positive tumors [39].

To identify direct mRNA targets of miR-21-3p, predicted targets from the MirTarBase, miRWalk and TargetScan were compared to the differentially expressed genes from the meta-analysis (**Figure 4a**). Our analysis revealed 129 potential targets of miR-21-3p that overlapped with genes in METABRIC and TCGA and were inversely correlated with miR-21-3p. Among these was PDCD4, a previously described miR-21-5p target.

Since each of the three databases suggested different direct targets of miR-21-3p, we focused on targets that were experimentally validated and limited our analysis to miR-21-3p targets from miRTarBase. According to this analysis, eight of the 70 genes were shared; and among these miR-21-3p inversely correlated with three: STARD13, MAT2A and ZNF132. STARD13 is a tumor suppressor that plays a role in breast cancer invasion and metastasis [40,41]. ZNF132 is implicated as a master transcriptional regulator of networks that underlie the breast cancer phenotype [42]. Additionally, high MAT2A expression predicts shorter distant metastasis-free survival in ER positive patients [43].

A network analysis identified downstream genes that might be regulated by miR-21-3p. The network edges are based on co-expression of genes from the BRCA-TCGA data analyzed using the ARACNe-AP algorithm [44]; and the nodes are selected from the lowest p-values in the meta-analysis. Collectively, these genes fall within a network that is important for cell proliferation, regulation of apoptosis, and cell migration (**Figure 4b**). Taken together, these results indicate miR-21-3p expression supports activation of pathways that facilitate tumor progression when their control is deregulated. Although our results are supported by some experimental evidence, further validation is needed.

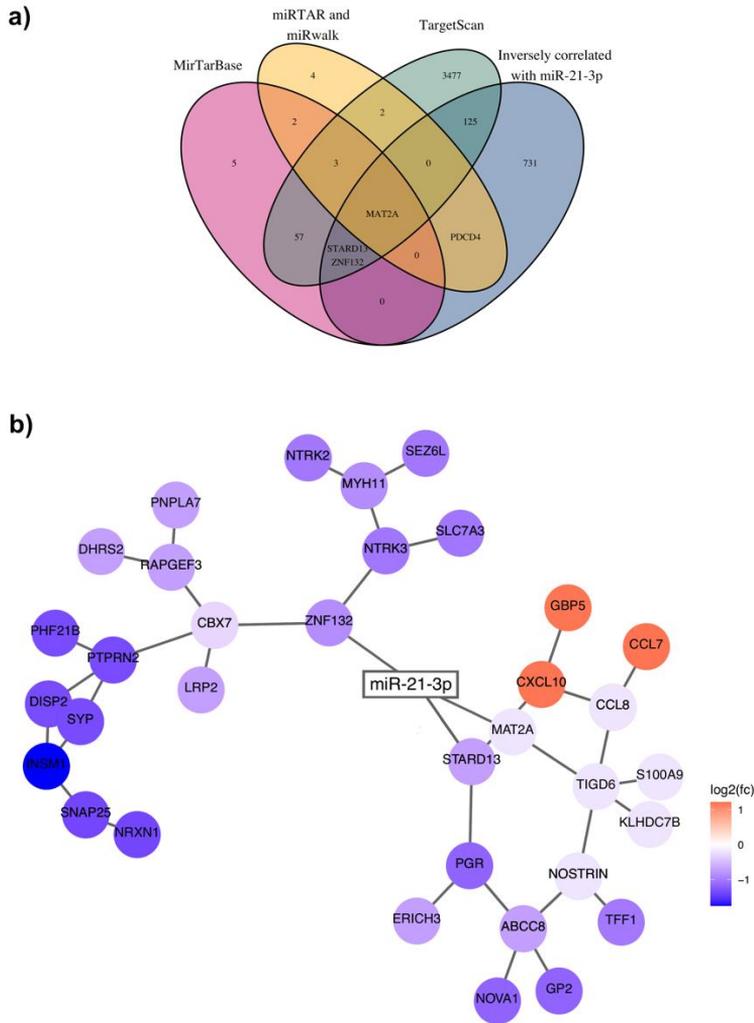


Figure 4. miR-21-3p validated targets viewed in a network. a) Venn diagram showing overlap between the genes identified in the analysis as inversely correlated with miR-21-3p (blue) and miRNA target databases miRTarBase (red), TargetScan (green), miRTAR and miRwalk (yellow). 129 genes from our analysis are listed as predicted targets of miR-21-3p in these databases out of which three, STARD13, ZNF132 and MAT2A, have been validated experimentally. b) Simplified gene co-expression network diagram showing the three validated targets of miR-21-3p. The network diagram was constructed using genes with expression that significantly correlated with miR-21-3p levels in the meta-analysis. The color represents the mean $\log_2(\text{FC})$. The edges between gene nodes were inferred by Aracne-AP algorithm and applied to breast cancer samples from TCGA.

4. Discussion

In our study, we demonstrated that high levels of miR-21-3p associated with pathoclinical characteristics of a worse prognosis and shorter BCSS. In addition, we verified that the miR-21-3p target genes, ZNF132, STARD13, and MAT2A, were significantly down-regulated when miR-21-3p expression was high. These genes are implicated in breast tumor development.

MiR-21 is up-regulated in many cancer types and can be up-regulated in several ways. It has its own promoter in intron 10 of VMP1, at 17q23.1 (see Fig. 1 in a review by Bautista-Sánchez et al. [6]). Its locus is frequently amplified in breast tumors, which can increase expression of the genes therein. If VMP1 is the 3' partner of a fusion gene, then mature miR-21 products increase [45]. VMP1-miR-21 fusion transcripts are known [2,4,43] that result in increased miR-21 products. The locus appears to be regulated by a complex regulatory mechanism, akin to the situation in colorectal cancer where an autoregulatory loop in between miR-21 and VMP1 [46] and a miR-21-3p isomiR is suggested to downregulate miR-21-5p [47]. MiR-21-3p reportedly is more abundant in tumors, including breast tumors [14,15], than in normal tissue, an effect we confirmed in the cohorts we analyzed.

Our results suggest elevated expression of miR-21-3p might serve as a prognostic marker in breast cancer. In the exploration cohort (cohort-1), miR-21-3p levels associated with shorter survival. Among the validation cohorts, high miR-21-3p levels also affected BCSS, with significance in the METABRIC cohort, a trend detected in cohort-2 but not in TCGA. This discrepancy between cohorts may be due to the treatments received by patients comprising each, despite that their clinical and pathological characteristics were similar they differed in their time at diagnosis, which in turn affects treatment. For example, patients in METABRIC and cohort-1 did not receive trastuzumab. In addition, the patients comprising each cohort had been treated with different drug combinations. Moreover, technical reasons might account for some differences, e.g., the techniques that measure miR-21-3p (see methods). Further complicating the analysis, a variety of miRs isoforms (called isomiRs) were identified in colorectal cancers [47]; and the probes we used to analyze cohorts 1 and 2 only captured the reference isomiR. In contrast, the METABRIC microarray probe is 15 nucleotides, and the probe sequence is embedded in the reference sequence. Since detection is based on hybridization, the METABRIC analysis has the potential to capture additional isomiRs. In TCGA, the miR-21-3p values used in this study represent nine isomiRs including the reference (**Figure S1**). These considerations are important as isomiRs affect the cellular transcriptome differently and can be differently expressed based on ethnicity [48].

The function of miR-21-3p may depend on tumor type and context. In non-small cell lung cancer (NSCLC) and esophageal squamous cell carcinoma (ESCC), miR-21-3p was more highly expressed in tumors than the adjacent normal tissue [49,50]. In ESCC, high miR-21-3p levels also associated with a high risk of cancer progression [50], in agreement with our findings. In cell-based assays in a colorectal cancer cell line [51] and in ovarian and prostate

cell lines [52], high miR-21-3p had oncogenic properties [51]; however, in a hepatocellular cancer cell line, miR-21-3p overexpression suppressed growth and increased apoptosis, suggesting tumor suppressive properties [53]. One breast cancer study focusing on miR-21-3p is in agreement with the results in the hepatocellular cancer cell line, as low levels of miR-21-3p were a risk factor for overall survival (OS) in patients with triple-negative breast tumors [14]. Another breast cancer study suggested that high miR-21-3p was a potential biomarker for early detection of breast cancer [16]. We did not analyze miR-21-3p expression levels with respect to early breast cancer in our study, but our data are in line with these results as they suggest that high miR-21-3p is a prognostic factor for BCSS. Depending on the cellular context, the particular miR-21-3p isomiR expressed [47], and the target gene(s), miR-21-3p may act as either a tumor suppressor or oncogene.

To gain insight into the biological role of miR-21-3p, we looked for genes with expression that increased with high miR-21-3p, and whether these genes associated with cellular pathways active in tumor progression. Unsurprisingly, some of the genes that most significantly correlated with high miR-21-3p were its genetic neighbors at 17q23.1. These genes are not confounders in the survival analysis, but their effect is difficult to separate from that of miR-21-3p. The significantly upregulated Hallmark pathways, such as the epithelial-to-mesenchymal transition, G2/M checkpoint control, and inflammatory response (**Figure 3b**) are well known to be active in cancer [54].

To identify direct targets of miR-21-3p, miRTarBase was used because it includes experimentally validated targets. The three targets—STARD13, ZNF132 and MAT2A—that were significantly down-regulated when miR-21-3p expression was high are genes already implicated in breast cancer [40-43]. STARD13 reportedly functions in cytoskeletal reorganization, proliferation, and motility, all of which are processes necessary for cancer progression [40]. Silencing of the transcription factor ZNF132 promotes progression in ESCC [55]; and its downregulation is associated with poor prognosis in prostate cancer [56], indicating tumor suppressive properties. MAT2A, Methionine Adenosyltransferase 2A, catalyzes the production of S-adenosylmethionine, which is important for most cellular processes. In contrast to ZNF132 and STARD13, upregulation of MAT2A is associated with poor prognosis [43].

A genetic network that expands from the three miR-21-3p targets we validated and are significantly downregulated includes genes implicated in processes known to affect cancer progression (**Figure 4b**). Among the genes in the ZNF132 node, CBX7 is implicated in cancer progression and EMT [57], PTPRN2 confers resistance to apoptosis [58], and NTRK3 is a receptor tyrosine kinase whose overactive kinase domain is implicated in growth and metastasis [59]. The STARD13 node includes the progesterone receptor PGR, which is interesting because high levels of its isoforms induce invasion and metastasis [60] and high PGR also indicates a good prognosis in ER-positive breast cancer [61]. ABCC8, a member of the MRP family involved in multidrug resistance, has a role in diabetes but recently expression patterns

of the ABC family were suggested to be new hallmarks of cancer [62]. CXCL10 is a chemokine that, among others, aids immune cells in infiltration of tumors [63]. Overexpression of NOSTRIN, in pancreatic cancer, suppresses migration and invasion [64]. The MAT2A node includes TIGD6, which has not been associated with cancer. Notably, most genes in the network have been linked to cancer progression.

Previous studies identified miR-21-3p target genes: in ovarian cells it targets NAV3 [65], a known tumor suppressor; in hepatocellular carcinoma, it targets SMAD7, an inhibitor of the TGF β pathway [66]; and in ESCC, it targets TRAF4 [50]. Conversely, miR-21-3p expression upregulates L1CAM, which promotes cell motility, invasion, metastasis, and chemoresistance [67]. L1CAM and NAV3 are not targets of miR-21-5p [65,67], but among other well-known targets of miR-21-5p [68], our study identified PDCD4 as being regulated by miR-21-3p as well. Although miR-21-3p might affect the same pathways as miR-21-5p, (e.g., invasion and metastasis), our data suggest it does so, at least in part, through targeting different genes. Indeed our data shows miR-21-3p is a prognostic marker in breast cancer independent of oncomiR miR-21-5p. These results highlight the importance of studying each strand of a mature miRNA (i.e., the 3p and the 5p), independently, to distinguish each component's biological function. In the case of miR-21 much effort has been put into studying miR-21-5p, yet our results indicate miR-21-3p also modulates breast cancer progression.

5. Conclusions

Using mRNA, miRNA, clinical, pathological and survival data from a selection of breast cancer patient cohorts, we identified miR-21-3p as a candidate prognostic marker for breast cancer that is associated with shorter breast cancer survival. It is inversely correlated with STARD13, ZNF132 and MAT2A, which are implicated in tumor development. Therefore, this interesting breast cancer candidate miR-21-3p warrants further investigation to fully understand its impact on breast cancer progression.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Figure S1:** Nine miR-21-3p isomiRs in the BRCA cohort from TCGA are significantly higher in cancer than in matched normal tissue, **Figure S2:** Breast cancer specific survival (BCSS) in cohort 2 and TCGA, **Figure S3:** MiR-21-3p levels were higher in breast tumors than paired normal breast tissues. **Figure S4:** MiR-21-3p expression levels correlated with PPM1D, VMP1 and RPS6KB1. **Table S1:** IsomiRs used from TCGA based on significantly higher expression in tumor compared to normal tissue, **Table S2:** Clinical and pathological characteristics of cohort-1, **Table S3:** Clinical and pathological characteristics of METABRIC BC cohort, **Table S4:** Clinical and pathological characteristics of cohort-2, **Table S5:** Clinical and pathological characteristics of TCGA BC cohort, **Table S6:** Meta-analysis from TCGA and METABRIC gene expression data, **Table S7:** GSEA from curated gene sets (c2) for genes that positively correlate with miR-21-3p, **Table S8:** GO analysis using R package enrichGO for genes inversely

correlated with miR-21-3p and GSEA analysis of GO genes sets for genes inversely correlated with miR-21-3p.

Author Contributions: Conceptualization, Arsalan Amirfallah and Inga Reynisdottir; Data curation, Arsalan Amirfallah, Hildur Knutsdottir and Inga Reynisdottir; Formal analysis, Arsalan Amirfallah, Hildur Knutsdottir and Inga Reynisdottir; Funding acquisition, Inga Reynisdottir; Methodology, Arsalan Amirfallah, Hildur Knutsdottir and Inga Reynisdottir; Project administration, Inga Reynisdottir; Resources, Adalgeir Arason, Oskar Johannsson, Bjarni Agnarsson and Rosa Barkardottir; Supervision, Inga Reynisdottir; Visualization, Arsalan Amirfallah, Hildur Knutsdottir and Inga Reynisdottir; Writing – original draft, Inga Reynisdottir; Writing – review & editing, Arsalan Amirfallah, Hildur Knutsdottir, Adalgeir Arason, Bylgja Hilmarsdottir, Oskar Johannsson, Bjarni Agnarsson, Rosa Barkardottir and Inga Reynisdottir. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants to IR, RBB, BAA, OTJ and AdAr from The Icelandic Centre for Research (grant number 152530-051, www.rannis.is), from The Scientific Fund of Landspítali – The National University Hospital in Iceland (grant numbers A-2016-033 and A-2019-042), The Scientific Fund of The Icelandic Cancer Society (<https://www.krabb.is/english/>) in 2017 and from Gongum saman (<https://gongumsaman.is/>) in 2013, 2017 and 2018.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of The National Bioethics Committee of Iceland (VSN-15-138, 27 October 2015).

Data: Publicly available datasets were analyzed in this study. The data from the European Genome-phenome archive can be found here: <https://ega-archive.org/> in study EGAS0000000122, miRNA dataset accession EGAD00010000438 and mRNA dataset accession EGAD00010000434. The METABRIC data was retrieved from the cBioPortal: <https://www.cbioportal.org/> under “invasive breast carcinoma.” Data generated by the TCGA research network can be accessed at: <https://www.cancer.gov/tcga>.

Acknowledgments: The authors would like to thank the following employees of the Department of Pathology: Edda S. Freysteinsdottir for extracting RNA and generating cDNA from the non-neoplastic breast tissue, preparing the miR-21-3p data for statistical analyses as well as preparing the graphical abstract, Gudrun Johannesdottir for quantifying miR-21-3p in breast tumors and non-neoplastic breast tissue from cohort-2, and Sigrun B. Krisjansdottir for the preparation of breast cancer patient tissue. The results shown in this study are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The graphical abstract was created with BioRender.com.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or

interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Stavast, C.J.; Erkeland, S.J. The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. *Cells* **2019**, *8*, doi:10.3390/cells8111465.
2. Ribas, J.; Ni, X.; Castanares, M.; Liu, M.M.; Esopi, D.; Yegnasubramanian, S.; Rodriguez, R.; Mendell, J.T.; Lupold, S.E. A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res* **2012**, *40*, 6821-6833, doi:10.1093/nar/gks308.
3. Cai, X.; Hagedorn, C.H.; Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **2004**, *10*, 1957-1966, doi:10.1261/rna.7135204.
4. Ribas, J.; Lupold, S.E. The transcriptional regulation of miR-21, its multiple transcripts, and their implication in prostate cancer. *Cell Cycle* **2010**, *9*, 923-929, doi:10.4161/cc.9.5.10930.
5. Fujita, S.; Ito, T.; Mizutani, T.; Minoguchi, S.; Yamamichi, N.; Sakurai, K.; Iba, H. miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *J Mol Biol* **2008**, *378*, 492-504, doi:10.1016/j.jmb.2008.03.015.
6. Bautista-Sanchez, D.; Arriaga-Canon, C.; Pedroza-Torres, A.; De La Rosa-Velazquez, I.A.; Gonzalez-Barrrios, R.; Contreras-Espinosa, L.; Montiel-Manriquez, R.; Castro-Hernandez, C.; Fragoso-Ontiveros, V.; Alvarez-Gomez, R.M.; et al. The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics. *Mol Ther Nucleic Acids* **2020**, *20*, 409-420, doi:10.1016/j.omtn.2020.03.003.
7. Ribas, J.; Lubold, S. The Role of miR-21, an Androgen-Responsive MicroRNA, in Prostate Cancer. *Androgen-Responsive Genes in Prostate Cancer: Regulation, Function and Clinical Applications* **2013**, 285 - 305, doi:10.1007/978-1-4614-6182-1.
8. Jinling, W.; Sijing, S.; Jie, Z.; Guinian, W. Prognostic value of circulating microRNA-21 for breast cancer: a systematic review and meta-analysis. *Artif Cells Nanomed Biotechnol* **2017**, *45*, 1-6, doi:10.1080/21691401.2016.1216856.
9. Binabaj, M.M.; Bahrami, A.; Khazaei, M.; Avan, A.; Ferns, G.A.; Soleimanpour, S.; Ryzhikov, M.; Hassanian, S.M. The Prognostic Value of Small Noncoding microRNA-21 Expression in the Survival of Cancer Patients: A Meta-Analysis. *Crit Rev Eukaryot Gene Expr* **2020**, *30*, 207-221, doi:10.1615/CritRevEukaryotGeneExpr.2020028719.
10. Nguyen, P.; Bar-Sela, G.; Sun, L.; Bisht, K.S.; Cui, H.; Kohn, E.; Feinberg, A.P.; Gius, D. BAT3 and SET1A form a complex with

CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression. *Mol Cell Biol* **2008**, *28*, 6720-6729.

11. Zhang, X.; Gee, H.; Rose, B.; Lee, C.S.; Clark, J.; Elliott, M.; Gamble, J.R.; Cairns, M.J.; Harris, A.; Khoury, S.; et al. Regulation of the tumour suppressor PDCD4 by miR-499 and miR-21 in oropharyngeal cancers. *BMC Cancer* **2016**, *16*, 86, doi:10.1186/s12885-016-2109-4.

12. Zhu, S.; Si, M.L.; Wu, H.; Mo, Y.Y. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J Biol Chem* **2007**, *282*, 14328-14336, doi:10.1074/jbc.M611393200.

13. Wang, X.S.; Prensner, J.R.; Chen, G.; Cao, Q.; Han, B.; Dhanasekaran, S.M.; Ponnala, R.; Cao, X.; Varambally, S.; Thomas, D.G.; et al. An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nat Biotechnol* **2009**, *27*, 1005-1011, doi:10.1038/nbt.1584.

14. Wu, X.; Ding, M.; Lin, J. Three-microRNA expression signature predicts survival in triple-negative breast cancer. *Oncol Lett* **2020**, *19*, 301-308, doi:10.3892/ol.2019.11118.

15. Ouyang, M.; Li, Y.; Ye, S.; Ma, J.; Lu, L.; Lv, W.; Chang, G.; Li, X.; Li, Q.; Wang, S.; et al. MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. *PLoS One* **2014**, *9*, e96228, doi:10.1371/journal.pone.0096228.

16. Yu, X.; Liang, J.; Xu, J.; Li, X.; Xing, S.; Li, H.; Liu, W.; Liu, D.; Xu, J.; Huang, L.; et al. Identification and Validation of Circulating MicroRNA Signatures for Breast Cancer Early Detection Based on Large Scale Tissue-Derived Data. *J Breast Cancer* **2018**, *21*, 363-370, doi:10.4048/jbc.2018.21.e56.

17. Fang, R.; Zhu, Y.; Hu, L.; Khadka, V.S.; Ai, J.; Zou, H.; Ju, D.; Jiang, B.; Deng, Y.; Hu, X. Plasma MicroRNA Pair Panels as Novel Biomarkers for Detection of Early Stage Breast Cancer. *Front Physiol* **2018**, *9*, 1879, doi:10.3389/fphys.2018.01879.

18. Amirfallah, A.; Arason, A.; Einarsson, H.; Gudmundsdottir, E.T.; Freysteinsdottir, E.S.; Olafsdottir, K.A.; Johannsson, O.T.; Agnarsson, B.A.; Barkardottir, R.B.; Reynisdottir, I. High expression of the vacuole membrane protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer. *PLoS One* **2019**, *14*, e0221413, doi:10.1371/journal.pone.0221413.

19. Gudmundsdottir, E.T.; Barkardottir, R.B.; Arason, A.; Gunnarsson, H.; Amundadottir, L.T.; Agnarsson, B.A.; Johannsson, O.T.; Reynisdottir, I. The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. *BMC Cancer* **2012**, *12*, 621.

20. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An

Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400-416 e411, doi:10.1016/j.cell.2018.02.052.

21. Curtis, C.; Shah, S.P.; Chin, S.F.; Turashvili, G.; Rueda, O.M.; Dunning, M.J.; Speed, D.; Lynch, A.G.; Samarajiwa, S.; Yuan, Y.; et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **2012**, *486*, 346-352.

22. Pereira, B.; Chin, S.F.; Rueda, O.M.; Vollan, H.K.; Provenzano, E.; Bardwell, H.A.; Pugh, M.; Jones, L.; Russell, R.; Sammut, S.J.; et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **2016**, *7*, 11479, doi:10.1038/ncomms11479.

23. Rueda, O.M.; Sammut, S.J.; Seoane, J.A.; Chin, S.F.; Caswell-Jin, J.L.; Callari, M.; Batra, R.; Pereira, B.; Bruna, A.; Ali, H.R.; et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* **2019**, *567*, 399-404, doi:10.1038/s41586-019-1007-8.

24. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2012**, *2*, 401-404, doi:10.1158/2159-8290.Cd-12-0095.

25. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **2013**, *6*, pl1, doi:10.1126/scisignal.2004088.

26. Dvinge, H.; Git, A.; Graf, S.; Salmon-Divon, M.; Curtis, C.; Sottoriva, A.; Zhao, Y.; Hirst, M.; Armitage, J.; Miska, E.A.; et al. The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature* **2013**, *497*, 378-382, doi:10.1038/nature12108.

27. Vire, E.; Curtis, C.; Davalos, V.; Git, A.; Robson, S.; Villanueva, A.; Vidal, A.; Barbieri, I.; Aparicio, S.; Esteller, M.; et al. The breast cancer oncogene EMSY represses transcription of antimetastatic microRNA miR-31. *Mol Cell* **2014**, *53*, 806-818, doi:10.1016/j.molcel.2014.01.029.

28. The R Project for Statistical Computing (<http://www.r-project.org>).

29. Jonsson, G.; Staaf, J.; Vallon-Christersson, J.; Ringner, M.; Holm, K.; Hegardt, C.; Gunnarsson, H.; Fagerholm, R.; Strand, C.; Agnarsson, B.A.; et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* **2010**, *12*, R42.

30. Andersen, C.L.; Monni, O.; Wagner, U.; Kononen, J.; Barlund, M.; Bucher, C.; Haas, P.; Nocito, A.; Bissig, H.; Sauter, G.; et al. High-throughput copy number analysis of 17q23 in 3520 tissue specimens by fluorescence in

situ hybridization to tissue microarrays. *Am J Pathol* **2002**, *161*, 73-79, doi:10.1016/S0002-9440(10)64158-2.

31. Inaki, K.; Hillmer, A.M.; Ukil, L.; Yao, F.; Woo, X.Y.; Vardy, L.A.; Zawack, K.F.; Lee, C.W.; Ariyaratne, P.N.; Chan, Y.S.; et al. Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome Res* **2011**, *21*, 676-687, doi:10.1101/gr.113225.110.

32. Haverty, P.M.; Fridlyand, J.; Li, L.; Getz, G.; Beroukhim, R.; Lohr, S.; Wu, T.D.; Cavet, G.; Zhang, Z.; Chant, J. High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer* **2008**, *47*, 530-542.

33. Natrajan, R.; Lambros, M.B.; Rodríguez-Pinilla, S.M.; Moreno-Bueno, G.; Tan, D.S.; Marchió, C.; Vatcheva, R.; Rayter, S.; Mahler-Araujo, B.; Fulford, L.G.; et al. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res* **2009**, *15*, 2711-2722, doi:10.1158/1078-0432.Ccr-08-1878.

34. Farmer, P.; Bonnefoi, H.; Becette, V.; Tubiana-Hulin, M.; Fumoleau, P.; Larsimont, D.; Macgrogan, G.; Bergh, J.; Cameron, D.; Goldstein, D.; et al. Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **2005**, *24*, 4660-4671, doi:10.1038/sj.onc.1208561.

35. Nikolsky, Y.; Sviridov, E.; Yao, J.; Dosymbekov, D.; Ustyansky, V.; Kaznacheev, V.; Dezso, Z.; Mulvey, L.; Macconail, L.E.; Winckler, W.; et al. Genome-wide functional synergy between amplified and mutated genes in human breast cancer. *Cancer Res* **2008**, *68*, 9532-9540, doi:10.1158/0008-5472.CAN-08-3082.

36. Higgins, M.; Obaidi, I.; McMorrow, T. Primary cilia and their role in cancer. *Oncol Lett* **2019**, *17*, 3041-3047, doi:10.3892/ol.2019.9942.

37. Weisbrod, A.B.; Zhang, L.; Jain, M.; Barak, S.; Quezado, M.M.; Kebebew, E. Altered PTEN, ATRX, CHGA, CHGB, and TP53 expression are associated with aggressive VHL-associated pancreatic neuroendocrine tumors. *Horm Cancer* **2013**, *4*, 165-175, doi:10.1007/s12672-013-0134-1.

38. Kimura, N.; Yoshida, R.; Shiraishi, S.; Pilichowska, M.; Ohuchi, N. Chromogranin A and chromogranin B in noninvasive and invasive breast carcinoma. *Endocr Pathol* **2002**, *13*, 117-122, doi:10.1385/ep:13:2:117.

39. Yoshida, R.; Ohuchi, N.; Kimura, N. Clinicopathological study of chromogranin A, B and BRCA1 expression in node-negative breast carcinoma. *Oncol Rep* **2002**, *9*, 1363-1367, doi:10.3892/or.9.6.1363.

40. Hanna, S.; Khalil, B.; Nasrallah, A.; Saykali, B.A.; Sobh, R.; Nasser, S.; El-Sibai, M. StarD13 is a tumor suppressor in breast cancer that regulates cell motility and invasion. *Int J Oncol* **2014**, *44*, 1499-1511, doi:10.3892/ijo.2014.2330.

41. Basak, P.; Leslie, H.; Dillon, R.L.; Muller, W.J.; Raouf, A.; Mowat, M.R.A. In vivo evidence supporting a metastasis suppressor role for Stard13

- (Dlc2) in ErbB2 (Neu) oncogene induced mouse mammary tumors. *Genes Chromosomes Cancer* **2018**, *57*, 182-191, doi:10.1002/gcc.22519.
42. Tovar, H.; Garcia-Herrera, R.; Espinal-Enriquez, J.; Hernandez-Lemus, E. Transcriptional master regulator analysis in breast cancer genetic networks. *Comput Biol Chem* **2015**, *59 Pt B*, 67-77, doi:10.1016/j.compbiolchem.2015.08.007.
43. Wang, C.Y.; Chiao, C.C.; Phan, N.N.; Li, C.Y.; Sun, Z.D.; Jiang, J.Z.; Hung, J.H.; Chen, Y.L.; Yen, M.C.; Weng, T.Y.; et al. Gene signatures and potential therapeutic targets of amino acid metabolism in estrogen receptor-positive breast cancer. *Am J Cancer Res* **2020**, *10*, 95-113.
44. Lachmann, A.; Giorgi, F.M.; Lopez, G.; Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **2016**, *32*, 2233-2235, doi:10.1093/bioinformatics/btw216.
45. Persson, H.; Søkilde, R.; Häkkinen, J.; Pirona, A.C.; Vallon-Christersson, J.; Kvist, A.; Mertens, F.; Borg, Å.; Mitelman, F.; Höglund, M.; et al. Frequent miRNA-convergent fusion gene events in breast cancer. *Nat Commun* **2017**, *8*, 788, doi:10.1038/s41467-017-01176-1.
46. Wang, C.; Peng, R.; Zeng, M.; Zhang, Z.; Liu, S.; Jiang, D.; Lu, Y.; Zou, F. An autoregulatory feedback loop of miR-21/VMP1 is responsible for the abnormal expression of miR-21 in colorectal cancer cells. *Cell Death Dis* **2020**, *11*, 1067, doi:10.1038/s41419-020-03265-4.
47. Jiao, W.; Leng, X.; Zhou, Q.; Wu, Y.; Sun, L.; Tan, Y.; Ni, H.; Dong, X.; Shen, T.; Liu, Y.; et al. Different miR-21-3p isoforms and their different features in colorectal cancer. *Int J Cancer* **2017**, *141*, 2103-2111, doi:10.1002/ijc.30902.
48. Telonis, A.G.; Loher, P.; Jing, Y.; Londin, E.; Rigoutsos, I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res* **2015**, *43*, 9158-9175, doi:10.1093/nar/gkv922.
49. Jiang, M.; Zhang, P.; Hu, G.; Xiao, Z.; Xu, F.; Zhong, T.; Huang, F.; Kuang, H.; Zhang, W. Relative expressions of miR-205-5p, miR-205-3p, and miR-21 in tissues and serum of non-small cell lung cancer patients. *Mol Cell Biochem* **2013**, *383*, 67-75, doi:10.1007/s11010-013-1755-y.
50. Gao, Z.; Liu, H.; Shi, Y.; Yin, L.; Zhu, Y.; Liu, R. Identification of Cancer Stem Cell Molecular Markers and Effects of hsa-miR-21-3p on Stemness in Esophageal Squamous Cell Carcinoma. *Cancers (Basel)* **2019**, *11*, doi:10.3390/cancers11040518.
51. Hou, N.; Guo, Z.; Zhao, G.; Jia, G.; Luo, B.; Shen, X.; Bai, Y. Inhibition of microRNA-21-3p suppresses proliferation as well as invasion and induces apoptosis by targeting RNA-binding protein with multiple splicing through Smad4/extra cellular signal-regulated protein kinase signalling

pathway in human colorectal cancer HCT116 cells. *Clin Exp Pharmacol Physiol* **2018**, *45*, 729-741, doi:10.1111/1440-1681.12931.

52. Baez-Vega, P.M.; Echevarria Vargas, I.M.; Valiyeva, F.; Encarnacion-Rosado, J.; Roman, A.; Flores, J.; Marcos-Martinez, M.J.; Vivas-Mejia, P.E. Targeting miR-21-3p inhibits proliferation and invasion of ovarian cancer cells. *Oncotarget* **2016**, *7*, 36321-36337, doi:10.18632/oncotarget.9216.

53. Lo, T.F.; Tsai, W.C.; Chen, S.T. MicroRNA-21-3p, a berberine-induced miRNA, directly down-regulates human methionine adenosyltransferases 2A and 2B and inhibits hepatoma cell growth. *PLoS One* **2013**, *8*, e75628, doi:10.1371/journal.pone.0075628.

54. Park, J.H.; Theodoratou, E.; Calin, G.A.; Shin, J.I. From cell biology to immunology: Controlling metastatic progression of cancer via microRNA regulatory networks. *Oncoimmunology* **2016**, *5*, e1230579, doi:10.1080/2162402X.2016.1230579.

55. Jiang, D.; He, Z.; Wang, C.; Zhou, Y.; Li, F.; Pu, W.; Zhang, X.; Feng, X.; Zhang, M.; Yecheng, X.; et al. Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma. *Cell Death Dis* **2018**, *10*, 1, doi:10.1038/s41419-018-1236-z.

56. Abildgaard, M.O.; Borre, M.; Mortensen, M.M.; Ulhoi, B.P.; Topping, N.; Wild, P.; Kristensen, H.; Mansilla, F.; Ottosen, P.D.; Dyrskjot, L.; et al. Downregulation of zinc finger protein 132 in prostate cancer is associated with aberrant promoter hypermethylation and poor prognosis. *Int J Cancer* **2012**, *130*, 885-895, doi:10.1002/ijc.26097.

57. Pallante, P.; Forzati, F.; Federico, A.; Arra, C.; Fusco, A. Polycomb protein family member CBX7 plays a critical role in cancer progression. *Am J Cancer Res* **2015**, *5*, 1594-1601.

58. Sorokin, A.V.; Nair, B.C.; Wei, Y.; Aziz, K.E.; Evdokimova, V.; Hung, M.C.; Chen, J. Aberrant Expression of proPTPRN2 in Cancer Cells Confers Resistance to Apoptosis. *Cancer Res* **2015**, *75*, 1846-1858, doi:10.1158/0008-5472.CAN-14-2718.

59. Friedrich, C.; Shalaby, T.; Oehler, C.; Pruschy, M.; Seifert, B.; Picard, D.; Remke, M.; Warmuth-Metz, M.; Kortmann, R.D.; Rutkowski, S.; et al. Tropomyosin receptor kinase C (TrkC) expression in medulloblastoma: relation to the molecular subgroups and impact on treatment response. *Childs Nerv Syst* **2017**, *33*, 1463-1471, doi:10.1007/s00381-017-3506-y.

60. Rosati, R.; Oppat, K.; Huang, Y.; Kim, S.; Ratnam, M. Clinical association of progesterone receptor isoform A with breast cancer metastasis consistent with its unique mechanistic role in preclinical models. *BMC Cancer* **2020**, *20*, 512, doi:10.1186/s12885-020-07002-0.

61. Kurozumi, S.; Matsumoto, H.; Hayashi, Y.; Tozuka, K.; Inoue, K.; Horiguchi, J.; Takeyoshi, I.; Oyama, T.; Kurosumi, M. Power of PgR

expression as a prognostic factor for ER-positive/HER2-negative breast cancer patients at intermediate risk classified by the Ki67 labeling index. *BMC Cancer* **2017**, *17*, 354, doi:10.1186/s12885-017-3331-4.

62. Dvorak, P.; Pesta, M.; Soucek, P. ABC gene expression profiles have clinical importance and possibly form a new hallmark of cancer. *Tumour Biol* **2017**, *39*, 1010428317699800, doi:10.1177/1010428317699800.

63. Wang, H.; Li, S.; Wang, Q.; Jin, Z.; Shao, W.; Gao, Y.; Li, L.; Lin, K.; Zhu, L.; Wang, H.; et al. Tumor immunological phenotype signature-based high-throughput screening for the discovery of combination immunotherapy compounds. *Sci Adv* **2021**, *7*, doi:10.1126/sciadv.abd7851.

64. Wang, J.; Yang, S.; He, P.; Schetter, A.J.; Gaedcke, J.; Ghadimi, B.M.; Ried, T.; Yfantis, H.G.; Lee, D.H.; Gaida, M.M.; et al. Endothelial Nitric Oxide Synthase Traffic Inducer (NOSTRIN) is a Negative Regulator of Disease Aggressiveness in Pancreatic Cancer. *Clin Cancer Res* **2016**, *22*, 5992-6001, doi:10.1158/1078-0432.CCR-16-0511.

65. Pink, R.C.; Samuel, P.; Massa, D.; Caley, D.P.; Brooks, S.A.; Carter, D.R. The passenger strand, miR-21-3p, plays a role in mediating cisplatin resistance in ovarian cancer cells. *Gynecol Oncol* **2015**, *137*, 143-151, doi:10.1016/j.ygyno.2014.12.042.

66. Hong, Y.; Ye, M.; Wang, F.; Fang, J.; Wang, C.; Luo, J.; Liu, J.; Liu, J.; Liu, L.; Zhao, Q.; et al. MiR-21-3p Promotes Hepatocellular Carcinoma Progression via SMAD7/YAP1 Regulation. *Front Oncol* **2021**, *11*, 642030, doi:10.3389/fonc.2021.642030.

67. Doberstein, K.; Bretz, N.P.; Schirmer, U.; Fiegl, H.; Blaheta, R.; Breunig, C.; Muller-Holzner, E.; Reimer, D.; Zeimet, A.G.; Altevogt, P. miR-21-3p is a positive regulator of L1CAM in several human carcinomas. *Cancer Lett* **2014**, *354*, 455-466, doi:10.1016/j.canlet.2014.08.020.

68. Buscaglia, L.E.; Li, Y. Apoptosis and the target genes of microRNA-21. *Chin J Cancer* **2011**, *30*, 371-380, doi:10.5732/cjc.011.10132.

Supplement Figures

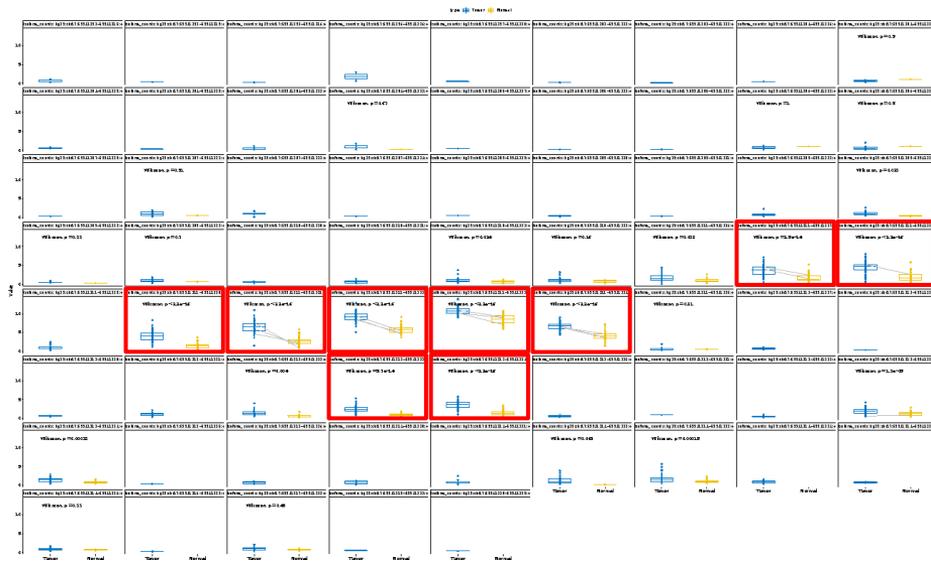


Figure S1. Nine miR-21-3p isomiRs in the BRCA cohort from TCGA are significantly higher in cancer than in matched normal tissue. Expression of all distinct miR-21-3p isomiRs in BRCA-TCGA data from matched tumor and normal samples. The isomiRs marked with a red box are the ones included in the analysis.

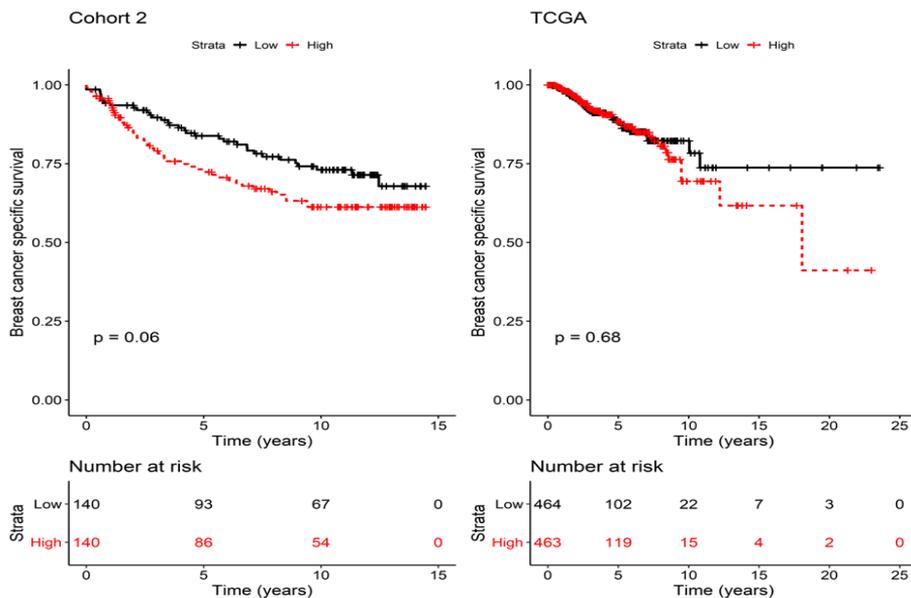


Figure S2. Breast cancer specific survival (BCSS) in cohort 2 and TCGA. Breast cancer specific survival (BCSS) was examined in (a) cohort-2 and (b)

TCGA. MiR-21-3p expression was used to divide the patients into two groups, high (red) and low (black) based on the median expression levels of miR-21-3p. The log rank p-values are shown in the graphs. The association of miR-21-3p with BCSS was borderline significant in (a) cohort-2 but not associated in (b) TCGA. The numbers of patients at risk at each time point are shown in tables below the graphs.

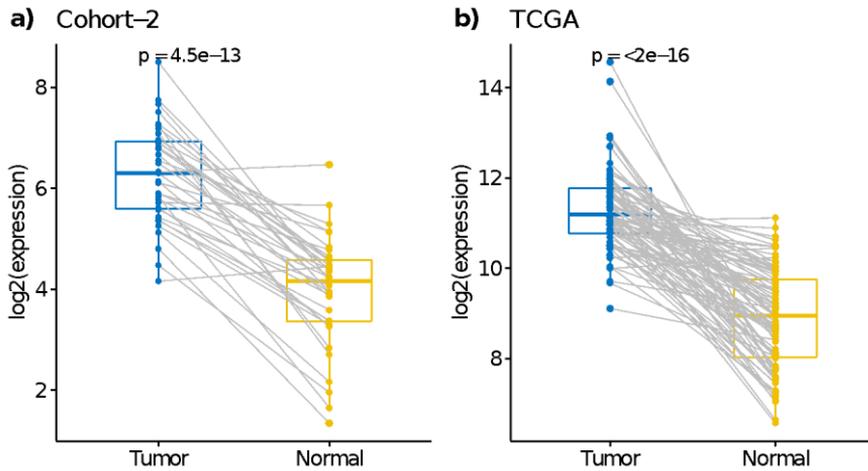


Figure S3. MiR-21-3p levels were higher in breast tumors than paired normal breast tissues. MiR-21-3p was examined in breast tumors and normal breast tissue from a) cohort-2 ($n = 35$) and b) TCGA ($n = 172$). A paired t-test was used to analyze expression levels between tumors and normal tissue. Expression in tumors was significantly higher than in normal breasts in cohort-2, $p = 4.5 \cdot 10^{-13}$ and TCGA, $p < 2 \cdot 10^{-16}$.

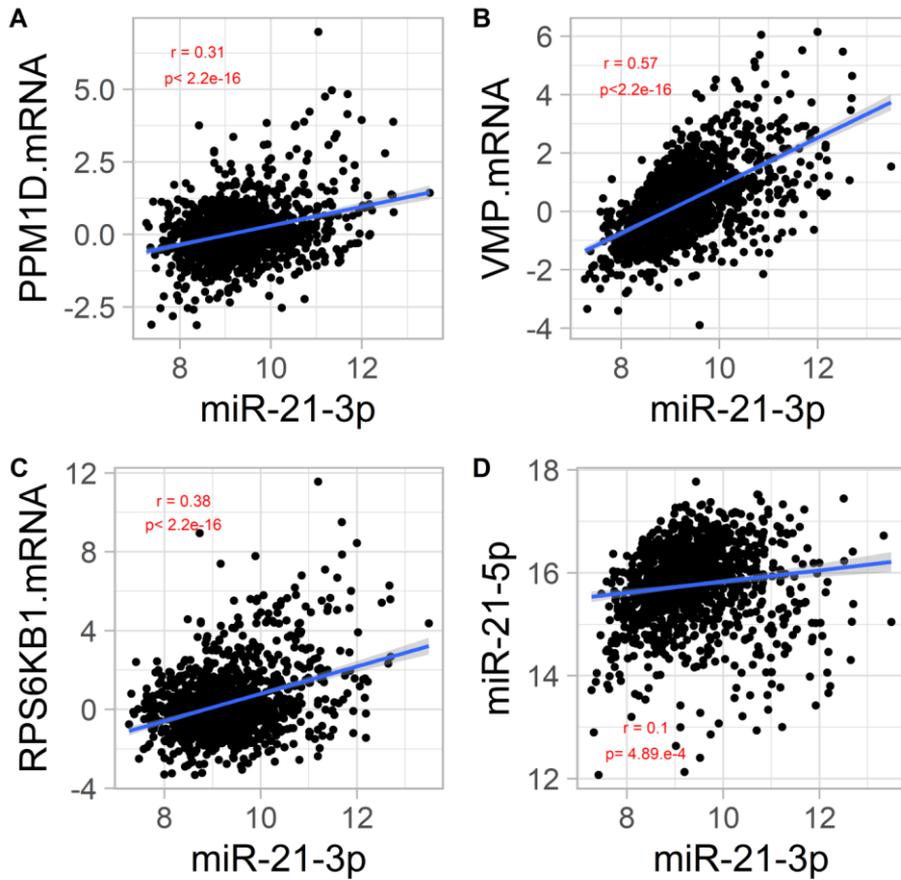


Figure S4. MiR-21-3p expression levels correlated with PPM1D, VMP1 and RPS6KB1. MiR-21-3p levels in METABRIC (measured via Agilent microarray) correlated with PPM1D mRNA, VMP1 mRNA and RPS6KB1 mRNA (measured by the same technique). (A) PPM1D and miR-21-3p Pearson r value was 0.31, $p < 2.2 \cdot 10^{-16}$; (B) VMP1 and miR-21-3p Pearson r value was 0.57, $p < 2.2 \cdot 10^{-16}$; (C) RPS6KB1 and miR-21-3p Pearson r value was 0.38, $p < 2.2 \cdot 10^{-16}$; (D) miR-21-5p and miR-21-3p Pearson r value was 0.11, $p = 4.89 \cdot 10^{-4}$.

Supplement Tables

Table s1. IsomiRs used from TCGA based on significantly higher expression in tumor compared to normal tissue

hg38:chr17:59841311-59841327:+
hg38:chr17:59841311-59841328:+
hg38:chr17:59841311-59841330:+
hg38:chr17:59841311-59841331:+
hg38:chr17:59841311-59841332:+
hg38:chr17:59841311-59841333:+
hg38:chr17:59841311-59841334:+
hg38:chr17:59841312-59841333:+
hg38:chr17:59841312-59841334:+

Table s2. Clinical and pathological characteristics of cohort 1

		miR21-3p mRNA	p-value
	13		
	9	median (25 and 75%)	
Age			0,263
< 50	53	0.114 (-0.420, 1.277)	
≥ 50	86	-0.157 (-1.007, 0.815)	
Estrogen receptor			0,483
Negative	46	-0.261 (-0.988, 0.818)	
Positive	89	0.116 (-0.677, 1.104)	
Unknown	4		
Progesterone receptor			0,361
Negative	63	-0.247 (-0.910, 0.801)	
Positive	69	0.1271 (-0.677, 1.201)	
Unknown	7		
HER2 status			0,286
	11		
Negative	6	-0.041 (-0.921, 1.025)	
Positive	22	0.239 (-0.330, 1.172)	

Unknown	1		
Tumor size			0,340
≤ 20	45	0.269 (-0.517, 1.018)	
> 20	94	-0.157 (-0.974, 0.997)	
Unknown	0		
Histologic Grade			0,473
1	12	0.130 (-0.439, 1.327)	
2	79	0.042 (-1.021, 1.032)	
3	47	-0.063 (-0.625, 0.774)	
Unknown	1		
Ki67			0,247
1	94	0.060 (-0.724, 1.071)	
2	28	-0.354 (-1.166, 0.395)	
3	15	-0.247 (-0.452, 0.750)	
Unknown	2		
Nodes			0,573
Negative	55	-0.063 (-0.862, 0.936)	
Positive	67	0.021 (-0.733, 1.186)	
Unknown	17		
Histology subtype			0,305
	12		
Ductal	0	-0.036 (-0.890, 0.905)	
Ductal_lobular	1	0,854	
Ductal_mucinous	1	0,713	
Lobular	11	0.429 (-0.069, 1.927)	
Medullary	4	-0.749 (-1.110, -0.061)	
Metaplastic	0	NA	
Metastasis_adenocarcinoma_breast_mucinous	1	(-0.997)	
Mucinous	1	(-1.136)	
Subtype Hu et al.			0,462
Basal	24	-0.361 (-0.975, 0.430)	
ERBB2	13	0.114 (-0.275, 0.817)	
LumA	42	0.159 (-0.902, 0.969)	
LumB	29	-0.059 (-1.161, 1.104)	
Normal	11	0.384 (-0.443, 2.594)	
unclassified	12	-0.053 (-0.694, 1.324)	
Unknown	6		

Metastasis			0,021
M0	62	-0.416 (-1.041, 0.691)	
M1	76	0.307 (-0.358, 1.306)	
Unknown	1		

Table s3. Clinical and pathological characteristics of METABRIC BC cohort

		miR21-3p mRNA	p-value
	1174	median (25 and 75%)	
Age			0,857
< 50	271	-0.004 (-0.474, 0.681)	
≥ 50	903	0.137 (-0.442, 0.550)	
Estrogen receptor			0,004
Negative	266	0.191 (-0.314, 0.719)	
Positive	908	-0.033 (-0.471, 0.528)	
Progesterone receptor			0,124
Negative	556	0.066 (-0.380, 0.644)	
Positive	618	-0.054 (-0.493, 0.524)	
HER2 status			2,63E-09
Negative	1026	-0.040 (-0.477, 0.496)	
Positive	148	0.459 (-0.088, 1.329)	
Tumor stage			0,007
0	1		
1	347	-0.082 (-0.509, 0.392)	
2	569	0.058 (-0.383, 0.653)	
3	96	0.023 (-0.321, 0.664)	
4	9	0.915 (0.187, 1.021)	
Unknown	152		
Tumor size			0,022
≤ 20	503	-0.018 (-0.456, 0.489)	
> 20	658	0.018 (-0.419, 0.675)	
Unknown	13		
Histologic Grade			3,68E-14
1	104	-0.252 (-0.516, 0.126)	

	2	469	-0.124 (-0.595, 0.314)	
	3	601	0.212 (-0.322, 0.839)	
Cellularity				0,016
	Low	130	-0.018 (-0.599, 0.520)	
	Moderate	430	-0.037 (-0.452, 0.525)	
	High	569	0.056 (-0.406, 0.697)	
	Unknown	45		
Nodes				0,001
	Negative	623	-0.082 (-0.518, 0.520)	
	Positive	551	0.061 (-0.348, 0.688)	
Nodal status				0,002
	N0	623	-0.082 (-0.518, 0.520)	
	N1	358	0.057 (-0.356, 0.564)	
	N2	127	0.234 (-0.367, 1.008)	
	N3	66	0.055 (-0.264, 0.807)	
Histology subtype				1,69E-07
	Ductal/NST	889	0.067 (-0.375, 0.682)	
	Lobular	92	-0.377 (-0.800, 0.024)	
	Medullary	15	-0.024 (-0.303, 0.729)	
	Metaplastic	0	NA	
	Mixed	138	-0.031 (-0.614, 0.304)	
	Mucinous	12	-0.646 (-0.970, -0.451)	
	Other	7	0.632 (-0.823, 1.737)	
	Tubular/cribriform	15	-0.288 (-0.692, 0.602)	
	Unknown	6		
Subtype PAM50				< 2E-16
	Basal	113	-0.082 (-0.489, 0.471)	
	Claudin-low	137	0.235 (-0.186, 0.718)	
	Her2	108	0.344 (-0.173, 1.045)	
	LumA	429	-0.145 (-0.564, 0.242)	
	LumB	285	0.226 (-0.342, 1.077)	
	NC	4	-0.154 (-0.571, 0.359)	
	Normal	98	-0.335 (-0.804, 0.457)	

3-Gene Classifier Subtype			< 2E-16
ER-/HER2-	185	0.065 (-0.352, 0.546)	
ER+/HER2- High Prolif	385	0.080 (-0.400, 0.793)	
ER+/HER2- Low Prolif	395	-0.161 (-0.602, 0.260)	
HER2+	122	0.483 (-0.015, 1.385)	
Unknown	87		

Table s4. Clinical and pathological characteristics of cohort-2

		miR21-3p mRNA	p-value
	281	median (25 and 75%)	
Age			0,965
< 50	64	0.014 (-0.659, 0.584)	
≥ 50	217	-0.009 (-0.554, 0.636)	
Estrogen receptor			0,036
Negative	77	0.191 (-0.448, 0.920)	
Positive	197	-0.037 (-0.730, 0.569)	
Unknown	7		
Progesterone receptor			0,024
Negative	94	0.194 (-0.456, 0.844)	
Positive	179	-0.092 (-0.781, 0.585)	
Unknown	8		
ERBB2 status			0,019
Negative	175	-0.080 (-0.786, 0.538)	
Positive	46	0.469 (-0.329, 1.317)	
Unknown	60		
HER2 status			0,003
0 (0 + 1)	217	-0.022 (-0.759, 0.551)	
1 (2 + 3)	51	0.398 (-0.341, 1.195)	
NA	13		
HER2 combo (ERBB2 status and HER2)			0,004
0	222	-0.024 (-0.721, 0.552)	
1	53	0.371 (-0.349, 1.171)	
NA	6		
Tumor size			0,078

≤ 20	83	-0.104 (-0.782, 0.546)	
> 20	196	0.016 (-0.528, 0.672)	
Unknown	2		
Histologic Grade			0,026
1	31	-0.080 (-0.841, 0.365)	
2	128	-0.083 (-0.737, 0.647)	
3	111	0.208 (-0.526, 0.744)	
Unknown	11		
Nodes			0,766
Negative	104	0.014 (-0.692, 0.749)	
Positive	146	0.004 (-0.551, 0.612)	
Unknown	31		
Histology subtype			0,457
Ductal	234	0.014 (-0.544, 0.712)	
Ductal_lobular	6	0.252 (-0.153, 0.553)	
Ductal_mixed	1	-0.759	
Lobular	29	-0.209 (-1.136, 0.420)	
Lobular_mixed	1	1,26	
Medullary	1	-0,042	
Metaplastic	2	0.788 (0.596, 0.979)	
Mucinous	5	0.022 (-1.454, 0.176)	
Sarcoma	1	(-0.328)	
Tubular	1	(-1.254)	

Table s5. Clinical and pathological characteristics of TCGA BC cohort

		miR21-3p mRNA	p-value
	946	median (25 and 75%)	
Age			0,053
< 50	270	0.050 (-0.366, 0.0533)	
≥ 50	676	-0.019 (-0.651, 0.504)	
Estrogen receptor			5,07E-05
Negative	216	0.227 (-0.273, 0.636)	
Positive	682	-0.075 (-0.682, 0.432)	
Unknown	48		
Progesterone receptor			8,69E-06

	Negative	296	0.196 (-0.370, 0.704)	
	Positive	599	-0.104 (-0.671, 0.403)	
	Unknown	51		
HER2 status				5,82E-04
	Negative	470	-0.045 (-0.655, 0.467)	
	Positive	131	0.177 (-0.317, 0.877)	
	Unknown	345		
Disease stage				0,994
	Stage I	167	0.010 (-0.626, 0.600)	
	Stage II	533	-0.019 (-0.547, 0.512)	
	Stage III	216	0.020 (-0.587, 0.482)	
	Stage IV	15	0.010 (-0.515, 0.719)	
	Stage X	10	0.122 (-0.455, 0.169)	
	Unknown	5		
Tumor stage				2,70E-02
	T1	258	0.008 (-0.546, 0.489)	
	T2	537	0.030 (-0.542, 0.553)	
	T3	118	-0.238 (-0.850, 0.306)	
	T4	30	0.211 (-0.299, 0.712)	
	TX	3	0.080 (-0.264, 0.589)	
Nodal status				0,195
	N0	441	-0.012 (-0.543, 0.540)	
	N1	319	-0.008 (-0.575, 0.449)	
	N2	106	0.060 (-0.433, 0.574)	
	N3	62	-0.144 (-0.924, 0.446)	
	NX	18	-0.321 (-0.591, 0.566)	
Nodes				0,553
	Negative	441	-0.012 (-0.543, 0.540)	
	Positive	505	0.005 (-0.603, 0.489)	
Metastasis				0,536
	Negative	776	-0.008 (-0.563, 0.534)	
	Positive	170	0.052 (-0.599, 0.426)	
Subtype PAM50*				<2e-16
	Basal-like	155	0.220 (-0.223, 0.179)	
	HER2-enriched	66	0.429 (-0.107, 0.981)	
	Luminal A	451	-0.150 (-0.8.2, 0.293)	
	Luminal B	160	0.237 (-0.234, 0.811)	

Normal-like	32	-0.469 (-1.057, 0.143)
Unknown	82	

Table s6. Meta-analysis from TCGA and METABRIC gene expression data

Meta-analysis from TCGA and METABRIC gene expression data						
gene	log2FoldC	pvalue.tcga	padj.deg.t	log2FoldC	pvalue.eg	padj.deg.e
VMP1	1,206606	7,71E-61	2,98E-57	1,029785	5,48E-107	1,05E-102
CHGB	-6,67252	2,94E-99	5,67E-95	-1,79786	3,40E-50	1,30E-46
PTRH2	1,129357	6,64E-42	6,42E-39	0,886846	3,73E-63	3,57E-59
PCSK1	-4,11152	3,16E-62	1,52E-58	-1,37247	9,00E-36	6,39E-33
CARTPT	-7,38541	2,14E-31	7,38E-29	-2,25632	1,39E-58	8,87E-55
PLAU	1,173516	1,76E-28	3,86E-26	1,096022	1,51E-57	7,22E-54
SNAP25	-2,65777	6,22E-39	4,51E-36	-1,62403	5,25E-44	1,01E-40
AQP9	2,042261	1,32E-33	5,41E-31	0,970967	4,47E-50	1,43E-46

Table s7. GSEA from curated gene sets (c2) for genes that positively correlate with miR-21-3p

geneset	link	pvalue
FARMER_BREAST_CANCER_CLUSTER_5	http://www.gsea-msigdb.org/gsea/msigdb/cards/FARMER_BREAST_CANCER_CLUSTER_5	2,2738E-23
NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON	http://www.gsea-msigdb.org/gsea/msigdb/cards/NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON	1,15688E-21
LASTOWSKA_NEUROBLASTOMA_COPY_NUMBER_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/LASTOWSKA_NEUROBLASTOMA_COPY_NUMBER_UP	2,36867E-17
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	4,44435E-17
POOLA_INVASIVE_BREAST_CANCER_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/POOLA_INVASIVE_BREAST_CANCER_UP	3,80527E-13
NABA_MATRISOME	http://www.gsea-msigdb.org/gsea/msigdb/cards/NABA_MATRISOME	5,60135E-13
PID_UPA_UPAR_PATHWAY	http://www.gsea-msigdb.org/gsea/msigdb/cards/PID_UPA_UPAR_PATHWAY	1,07767E-11
PID_INTEGRIN1_PATHWAY	http://www.gsea-msigdb.org/gsea/msigdb/cards/PID_INTEGRIN1_PATHWAY	1,78688E-10
PID_INTEGRIN3_PATHWAY	http://www.gsea-msigdb.org/gsea/msigdb/cards/PID_INTEGRIN3_PATHWAY	3,2474E-10
NABA_MATRISOME_ASSOCIATED	http://www.gsea-msigdb.org/gsea/msigdb/cards/NABA_MATRISOME_ASSOCIATED	1,41955E-09
RUTELLA_RESPONSE_TO_HGF_VS_CSF2RB_AND_IL4_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/RUTELLA_RESPONSE_TO_HGF_VS_CSF2RB_AND_IL4_UP	3,35989E-09
ZIRN_TRETINOIN_RESPONSE_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/ZIRN_TRETINOIN_RESPONSE_UP	3,83258E-09
GUILLAUMOND_KLF10_TARGETS_DN	http://www.gsea-msigdb.org/gsea/msigdb/cards/GUILLAUMOND_KLF10_TARGETS_DN	6,04248E-09
PID_FRA_PATHWAY	http://www.gsea-msigdb.org/gsea/msigdb/cards/PID_FRA_PATHWAY	7,55203E-09
MEBARKI_HCC_PROGENITOR_WNT_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/MEBARKI_HCC_PROGENITOR_WNT_UP	8,45652E-09
MCLACHLAN_DENTAL_CARIES_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/MCLACHLAN_DENTAL_CARIES_UP	9,62854E-09
CROMER_TUMORIGENESIS_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/CROMER_TUMORIGENESIS_UP	1,4783E-08
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_UP	2,19381E-08
PID_INTEGRIN2_PATHWAY	http://www.gsea-msigdb.org/gsea/msigdb/cards/PID_INTEGRIN2_PATHWAY	2,47642E-08

RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_DN	http://www.gsea-msigdb.org/gsea/msigdb/cards/RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_DN	6,11726E-08
WP_COMPLEMENT_AND_COAGULATION_CASCADES	http://www.gsea-msigdb.org/gsea/msigdb/cards/WP_COMPLEMENT_AND_COAGULATION_CASCADES	6,65905E-08
VERHAAK_AML_WITH_NPM1_MUTATED_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/VERHAAK_AML_WITH_NPM1_MUTATED_UP	7,25091E-08
VERHAAK_GLIOMASTOMA_MESENCHYMAL	http://www.gsea-msigdb.org/gsea/msigdb/cards/VERHAAK_GLIOMASTOMA_MESENCHYMAL	8,85125E-08
NABA_ECM_REGULATORS	http://www.gsea-msigdb.org/gsea/msigdb/cards/NABA_ECM_REGULATORS	1,16778E-07
SMIRNOV_CIRCULATING_ENDOTHELIOCYTES_IN_CANCER_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/SMIRNOV_CIRCULATING_ENDOTHELIOCYTES_IN_CANCER_UP	1,64057E-07
ANASTASSIOU_MULTICANCER_INVASIVENESS_SIGNATURE	http://www.gsea-msigdb.org/gsea/msigdb/cards/ANASTASSIOU_MULTICANCER_INVASIVENESS_SIGNATURE	1,78858E-07
WINZEN_DEGRADED_VIA_KHSRP	http://www.gsea-msigdb.org/gsea/msigdb/cards/WINZEN_DEGRADED_VIA_KHSRP	2,30744E-07
DELYS_THYROID_CANCER_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/DELYS_THYROID_CANCER_UP	3,19378E-07
ZHU_SKIL_TARGETS_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/ZHU_SKIL_TARGETS_UP	3,30239E-07
VART_KSHV_INFECTION_ANGIOGENIC_MARKERS_UP	http://www.gsea-msigdb.org/gsea/msigdb/cards/VART_KSHV_INFECTION_ANGIOGENIC_MARKERS_UP	4,53852E-07
XU_HGF_SIGNALING_NOT_VIA_AKT1_6HR	http://www.gsea-msigdb.org/gsea/msigdb/cards/XU_HGF_SIGNALING_NOT_VIA_AKT1_6HR	7,0048E-07
PETROVA_ENDOTHELIUM_LYMPHATIC_VS_BLOOD_DN	http://www.gsea-msigdb.org/gsea/msigdb/cards/PETROVA_ENDOTHELIUM_LYMPHATIC_VS_BLOOD_DN	7,55449E-07

GO analysis using R package enrichGO for genes inversely correlated with

ID	Descriptio	GeneRatic	BgRatio	pvalue	p.adjust	qvalue
GO:006007	synapse n	8/723	24/15013	1,03E-05	0,056921	0,055877
GO:006027	cilium ass	34/723	333/15013	2,97E-05	0,068898	0,067635
GO:004478	cilium org	35/723	351/15013	3,74E-05	0,068898	0,067635
GO:003606	ciliary bas	19/761	155/19559	9,71E-06	0,006368	0,005784
GO:007003	synaptobr	3/761	4/19559	0,000228	0,074746	0,067886
GO:003099	intraciliar	6/761	27/19559	0,000499	0,08748	0,07945
GO:003067	synaptic v	12/761	104/19559	0,000719	0,08748	0,07945
GO:009956	exocytic v	12/761	104/19559	0,000719	0,08748	0,07945

Appendix

Appendix

1. Characteristics of patients in cohort 1, cohort 2, TCGA, METABRIC and METABRIC/EGA cohorts.

Appendix Table 1. Patient characteristics of cohort 1

Character	n=158 (%)
Age	56 (27-88)
Estrogen receptor	
positive	100 (63.29)
negative	53 (33.54)
unknown	5 (3.16)
Progesterone receptor	
positive	82 (51.89)
negative	68 (43.03)
unknown	8 (5.06)
HER2 status	
positive	27 (17.08)
negative	130 (82.27)
unknown	1 (0.63)
Receptors ER and HER2	
ER- and HER2-	37 (23.41)
ER- and HER2+	15 (9.49)
ER+ and HER2-	89 (56.32)
ER+ and HER2+	11 (6.96)
unknown	6 (3.79)
Tumor size (mm)	25 (9-120)
Histological type	
IDC(a)	134 (84.81)
ILC(b)	12 (7.59)
other	12 (7.59)
Nodal status	
positive	81 (51.26)

negative	59 (37.34)
unknown	18 (11.39)
Ki 67	
High	45 (28.48)
Low	110 (69.62)
Unknown	3 (1.89)
Histological grade	
1	16 (10.12)
2	88 (55.69)
3	53 (33.54)
unknown	1 (0.63)
Metastasis	
Positive	68 (43.03)
Negative	89 (56.32)
unknown	1 (0.63)
Intrinsic subtype Hu et al.	
Basal	25 (15.82)
ERBB2	16 (10.12)
Luminal A	48 (30.37)
Luminal B	31 (19.62)
Normal-like	16 (10.12)
unknown	22 (13.92)
Familial status	
BRCA2	113 (71.51)
Non-BRCA2	43 (27.21)

(a)IDC: Invasive ductal tumors.

(b)ILC: Invasive lobular tumors.

Appendix Table 2. Patient characteristics of cohort 2

Characteristic	n=277 (%)
Age (year)	
median (range)	60 (27-97)
Estrogen receptor	
positive	194 (70)
negative	77 (28)
unknown	6 (2)
Progesterone receptor	
positive	176 (63.5)
negative	94 (34)
unknown	7 (2.5)
HER2 status	
positive	55 (20)
negative	217 (78)
unknown	5 (2)
Tumor size (mm)	
>20 mm	194 (70)
≤20 mm	82 (29.6)
unknown	1 (0.4)
Histological type	
IDC	231(83)
ILC	30 (11)
other	16 (6)
Nodal status	
positive	146 (53)
negative	101 (36)
unknown	30 (11)
Histological grade	
1	31 (11)
2	124 (45)
3	107 (39)
unknown	15 (5)
Metastasis	
positive	65 (23.5)
negative	210 (75.8)
unknown	2 (0.7)

Appendix Table 3. Patient characteristics of TCGA

Characteristic	n=816 (%)
Age at breast cancer diagnosis	
Median (range)	59 (26 - 90)
Tumor Stage N (%)	
1	219 (10.04)
2	458 (56.12)
3	105 (12.86)
4	3 (0.36)
NA	31 (3.79)
Nodes	
Negative	382 (46.81)
Positive	420 (51.47)
NA	14 (1.71)
Histologic Subtype	
Ductal/NST	598 (73.28)
Lobular	143 (17.52)
Medullary	5 (0.61)
Metaplastic	3 (0.36)
Mixed	23 (2.81)
Mucinous	14 (1.71)
Other	28 (3.43)
NA	2 (0.24)
Progesterone	
Positive	521 (63.84)
Negative	251 (30.75)
NA	44 (3.59)
Estrogen	
Positive	600 (73.52)
Negative	175 (21.44)
NA	41 (5.02)
HER2	
Positive	416 (50.98)
Negative	121 (14.82)
Na	279 (34.19)

Subtype PAM50 N (%)

Luminal A	414 (50.73)
Luminal B	176 (21.56)
HER2 enriched	65 (70.96)
Normal like	24 (2.94)
NA	137 (16.78)

Appendix Table 4. Patient characteristics of METABRIC

Characteristic	n=2509 (%)
Age at breast cancer diagnosis	
Median (range)	61 (22 - 96)
Tumor Stage N (%)	
0	24 (1.3)
1	630 (35.2)
2	979 (54.8)
3	144 (8.1)
4	11 (0.6)
NA	721 (0)
Nodes status	
N0	1196 (53.3)
N1	695 (31.0)
N2	233 (10.4)
N3	119 (5.3)
NA	266
Nodes	
Negative	1196 (53.3)
Positive	1047 (46.7)
NA	266
Histologic Subtype	
Ductal/NST	1810 (76.2)
Lobular	192 (8.1)
Medullary	32 (1.3)
Metaplastic	2 (0.1)
Mixed	269 (11.3)
Mucinous	25 (1.1)
Other Tubular	21 (0.9)
Cribriform	23 (1.0)
NA	135

Tumor Size (mm)	
Median (range)	26.22 (15.37)
Neoplasm Histologic Grade	
1	214 (8.9)
2	976 (40.9)
3	1198 (50.2)
NA	121
Cellularity	
Low	215 (11.2)
Moderate	737 (38.4)
High	965 (50.3)
NA	592
Histopathology N (%)	
ER+	1825 (73.9)
PR+	1040 (52.5)
HER2+	247 (12.5)
Triple negative	320 (16.2)
Molecular subtype	
Subtype PAM50 N (%)	
Basal-like	209 (10.6)
Claudin-low	218 (11.0)
HER2	224 (11.3)
Luminal A	700 (35.4)
Luminal B	475 (24)
NC	6 (0.3)
Normal-like	148 (7.5)
NA	529
Three-gene classifier subtype	
ER-/HER2-	309 (17.5)
ER+/HER2- High Prolif	617 (35.0)
ER+/HER2- Low Prolif	640 (36.3)
HER2+	198 (11.2)
NA	745

Appendix Table 5. Patient characteristics of METABRIC/EGA

Characteristic	n=1220 (%)
Age at breast cancer diagnosis	
Median (range)	61 (22 - 96)
Tumor Stage N (%)	
0	2 (0.2)
1	364 (34.3)
2	588 (55.4)
3	98 (9.2)
4	10 (9.0)
NA	158
Node status	
N0	623 (53.1)
N1	358 (30.5)
N2	127 (10.8)
N3	66 (5.6)
NA	46
Nodes	
Negative	623 (53.1)
Positive	551 (46.9)
NA	46
Histologic Subtype	
Ductal/NST	922 (75.7)
Lobular	96 (7.9)
Medullary	15 (1.2)
Metaplastic	139 (11.4)
Mixed	12 (1.0)
Mucinous	12 (1.0)
Other Tubular	7 (0.6)
Cribriform	15 (1.2)
NA	2 (0.2)
Tumor Size (mm)	
Mean (SD)	26.36 (15.71)
Neoplasm Histologic Grade	
1	106 (8.7)

2	494 (40.5)
3	620 (50.8)
NA	0
Cellularity	
Low	137 (11.7)
Moderate	447 (38.2)
High	586 (50.1)
NA	50
Histopathology N (%)	
ER+	939 (77.0)
PR+	639 (52.4)
HER2+	153 (12.5)
Triple negative	197 (16.1)
Subtype PAM50 N (%)	
Basal-like	118 (10.2)
Claudin-low	148 (12.8)
HER2	109 (9.4)
Luminal A	385 (33.2)
Luminal B	294 (25.3)
NC	4 (0.3)
Normal-like	102 (8.8)
NA	60
Three-gene classifier subtype	
ER-/HER2-	196 (17.4)
ER+/HER2- High Prolif	394 (34.9)
ER+/HER2- Low Prolif	411 (36.4)
HER2+	127 (11.3)
NA	92

2. Characteristic of cell lines and list of fusion genes used in *in-silico* analysis.

Appendix Table 6. Fusion genes from breast cancer cell lines that were used in the *in-silico* analysis

Gene	Gene B2	cell line	source
NOTCH1	GABBR2	BT20	PMC3233654
GOLGB1	ILDR1	BT20	PMC3233654
PLEKHB2	ARHGEF4	BT20	PMC3233654
LIMA1	USP22	BT20	PMC3159479
VAPB	IKZF3	BT474	PMC3091304
RAB22A	MYO9B	BT474	PMC3091304
SKA2	MYO19	BT474	PMC3091304
DIDO1	KIAA0406	BT474	PMC3091304
STARD3	DOK5	BT474	PMC3091304
LAMP1	MCF2L	BT474	PMC3091304
GLB1	CMTM7	BT474	PMC3091304
CPNE1	PI3	BT474	PMC3091304
ZMYND8	CEP250	BT474	PMC3159479
MED1	STXBP4	BT474	PMC3159479 / SOAPfuse/ PMC3485361 PMC3159479
PIP4K2B	RAD51C	BT474	/PMC3431177/PMC3485361
MED13	BCAS3	BT474	PMC3431177
NCOA2	ZNF704	BT474	PMC3431177
MYO9B	FCHO1 9	BT474	PMC3431177
STX16	RAE1	BT474	PMC3431177
RPS6KB1	SNF8	BT474	PMC3431177/PMC3159479 /SOAPfuse
USP32	MED1	BT474	PMC3485361
THRA	AC090627.1	BT474	PMC3485361
ZMYND8	CEP250	BT474	SOAPfuse
ACACA	STAC2	BT474	SOAPfuse
STX16	RAE1	BT474	SOAPfuse
TOB1	SYNRG	BT474	SOAPfuse/PMC3431177
MED1	ACSF2	BT474	SOAPfuse/PMC3485361
THRA	AC090627.1	BT474	PMID: 25485619 PMC3159479 / SOAPfuse/ PMC3431177
TRPC4AP	MRPL45	BT474	PMC3431177
SMARCB1	MARK3	BT483	PMC3431177

1

CLTC	VMP1	BT549	PMC3431177
ST7	PRKAG2	CAMA1	PMC3233654
PLDN	SQRDL	CAMA1	PMC3233654
FBRS	ZNF771	EFM19	PMC3233654
ZFYVE9	USP33	EFM19	PMC3233654
RFX1	ASNA1	HCC1008	PMC3233654
C18orf45	HM13	HCC1143	PMC3431177
C2ORF48	RRM2	HCC1143	PMC3431177
CTAGE5	SIP1	HCC1187	PMC3159479
PUM1	TRERF1	HCC1187	PMC3431177
SEC22B	NOTCH2	HCC1187	PMC3431177
AGPAT5	MCPH1	HCC1187	PMC3431177
PLXND1	TMCC1	HCC1187	PMID: 25485619
RGS22	SYCP1	HCC1187	PMID: 25485619
CTAGE5	GEMIN2	HCC1187	PMID: 25485619
SUSD1	PTBP3	HCC1187	PMID: 25485619
EIF3K	CYP39A1	HCC1395	PMC3159479
RAB7A	LRCH3	HCC1395	PMC3159479
HNRNPU L2	AHNAK 11	HCC1395	PMC3431177
EFTUD2	KIF18B	HCC1395	PMID: 25485619
ERO1L	FERMT2	HCC1395	PMID: 25485619
KCNQ5	RIMS1	HCC1395	PMID: 25485619
PLA2R1	RBMS1	HCC1395	PMID: 25485619
PLEC1	C8ORF38	HCC1419	PMC3233654
SLC37A1	ABCG1	HCC1428	PMC3159479
RNF187	OBSCN	HCC1428	PMC3159479
PPP1R1B	STARD3	HCC1569	PMC3233654
PSD3	CHGN	HCC1569	PMC3233654
CYTH1	PRPSAP1	HCC1599	PMC3159479
PSCD1	PRPSAP1	HCC1599	PMC3233654
EXOC7	CYTH1	HCC1599	PMC3159479
TAX1BP1	AHCY	HCC1806	PMC3159479
BRE	DPYSL5	HCC1806	PMC3159479
CD151	DRD4	HCC1806	PMC3159479
LDLRAD3	TCP11L1	HCC1806	PMC3159479
RFT1	UQCRC2	HCC1806	PMC3159479
NFIA	EHF	HCC1937	PMC3159479
INTS1	PRKAR1B	HCC1954	PMC3159479

GSDMC	PVT1	HCC1954	PMC3159479
INTS1	PRKAR1B	HCC1954	PMC3233654
C6orf106	SPDEF	HCC1954	PMC3431177
STRADB	NOP58	HCC1954	PMC3159479
PHF20L1	SAMD12	HCC1954	PMC3159479
FBXL20	SNF8	HCC202	PMC3233654
RASA2	ACPL2	HCC2157	PMID: 25485619
SMYD3	ZNF695	HCC2157	PMID: 25485619
POLDIP2	BRIP1	HCC2218	PMC3159479
SEC16A	NOTCH1	HCC2218	PMC3431177
PERLD1	PPM1D	HCC2218	PMC3431177
BCL2L12	PRMT1	HCC38	PMC3233654
ACBD6	RRP15	HCC38	PMID: 25485619
LDHC	SERGEF	HCC38	PMID: 25485619
MBOAT2	PRKCE	HCC38	PMID: 25485619
SLC26A6	PRKAR2A	HCC38	PMID: 25485619
HMGXB3	PPARGC1B	HCC38	PMID: 25485619
BSG	NFIX	KPL4	PMID: 25485619
PPP1R12 A	SEP10(0)	KPL4	PMID: 25485619
NOTCH1	NUP214	KPL4	PMID: 25485619
SULF2	PRICKLE2	MCF7	PMC3091304/PMC3431177/PMC3159479
ATXN7L3	FAM171A2	MCF7	PMC3159479
RPS6KB1	DIAPH3	MCF7	PMC3159479 / SOAPfuse
PAPOLA	AK7	MCF7	PMC3431177
AHCYL1	RAD51C	MCF7	PMC3431177
ARHGAP1 9	DRG1	MCF7	PMC3431177
HSPE1	PREI3	MCF7	PMC3431177
TRIM37	VMP1	MCF7	PMC3431177
BCAS4	ZMYND8	MCF7	PMC3431177
PVT1	MYC	MCF7	PMC3431177
RPS6KB1 AC099850 .1	VMP1	MCF7	PMC3431177/PMC3091304/SOAPfuse
MYH9	EIF3D	MCF7	PMC3485361
CRIP2	CRIP1	MCF7	SOAPfuse
BCAS4	BCAS3	MCF7	SOAPfuse
ARFGEF2	SULF2	MCF7	SOAPfuse

ATP1A1	ZFP64	MCF7	SOAPfuse
DEPDC1B	ELOVL7	MCF7	SOAPfuse
GATAD2B	NUP210L	MCF7	SOAPfuse
MYO6	SENP6	MCF7	SOAPfuse
POLA2	CDC42EP2	MCF7	SOAPfuse
POP1	MATN2	MCF7	SOAPfuse
SMARCA4	CARM1	MCF7	SOAPfuse/ PMC3485361
GCN1L1	MSI1	MCF7	SOAPfuse/ PMC3485361
RAD51C	ATXN7	MCF7	PMID: 25485619
SLC25A24	NBPF6	MCF7	PMID: 25485619
USP31	CRYL1	MCF7	PMID: 25485619
TBL1XR1	RGS17	MCF7	PMID: 25485619
TAF4	BRIP1	MCF7	PMID: 25485619
ABCA5	PPP4R1L	MCF7	PMID: 25485619
C16orf45	ABCC1	MCF7	PMID: 25485619
C16orf62	IQCK	MCF7	PMID: 25485619
CXorf15	SYAP1	MCF7	PMID: 25485619
SYTL2	PICALM	MCF7	PMID: 25485619
BC035340	MCF2L	MDAMB134 MDAMB134	PMC3233654 MediSapiens (www.medisapiens.com)
ANK1	ZMAT4	VI	
CCDC9	KIAA0134	MDAMB157	PMC3233654
TYRO3	RTF1	MDAMB157 MDAMB175	PMC3233654
ODZ4	NRG1	VII	PMC3159479
ANKHD1	CYSTM1	MDAMB231	SOAPfuse/MediSapiens (www.medisapiens.com)
ANKHD1	PCDH1	MDAMB231	SOAPfuse/MediSapiens (www.medisapiens.com)
ANKHD1-EIF4EBP3	PCDH1	MDAMB231	SOAPfuse/MediSapiens (www.medisapiens.com)
BRIP1	VMP1	MDAMB361	PMC3159479
SUPT4H1	CCDC46	MDAMB361	PMC3159479
TMEM104	CDK12	MDAMB361	PMC3159479
TMEM104	CRKRS	MDAMB361	PMC3431177
TANC1	MTMR4	MDAMB361	PMC3431177
RIMS2	ATP6V1C1	MDAMB436	PMC3159479
TIAL1	C10orf119	MDAMB436	PMC3159479
MECP2	TMLHE	MDAMB453	PMC3159479
ARID1A	MAST2	MDAMB468	PMC3431177
UBR5	SLC25A32	MDAMB468	SOAPfuse/PMC3431177/PMC3159

TATDN1	GSDMB ENSG00000236	SKBR3	PMC3091304
CSE1L	127	SKBR3	PMC3091304
RARA	PKIA	SKBR3	PMC3091304
CCDC85C	SETD3	SKBR3	PMC3091304
SUMF1	LRRFIP2	SKBR3	PMC3091304
WDR67	ZNF704	SKBR3	PMC3091304
CYTH1	EIF3H	SKBR3	PMC3091304
DHX35	ITCH	SKBR3	PMC3091304
NFS1	PREX1	SKBR3	PMC3091304
KLHDC2	SNTB1	SKBR3	PMC3159479
RAB43P1	CNBP	SUM225	SOAPfuse
CLIC4	FAM132A	SUM225	SOAPfuse/MediSapiens (www.medisapiens.com)
RAB43	CNBP	SUM225	SOAPfuse/MediSapiens (www.medisapiens.com)
FGFR2	ACADSB	SUM52	MediSapiens (www.medisapiens.com)
SUPT5H	SIPA1L3	SUM52	SOAPfuse
VGLL4	SH3BP5	T47D	SOAPfuse
ARID1A	WDTC1	UACC812	PMC3159479
HDGF	S100A10	UACC812	PMC3159479
PPP1R12 B	SNX27	UACC812	PMC3159479
WIPF2	ERBB2	UACC812	PMC3159479
SRGAP2	PRPF3	UACC812	PMC3159479
CDC6	IKZF3	UACC812	PMC3233654
MLLT6	TEM7	UACC812	PMC3233654
CCDC6	ANK3	UACC893	MediSapiens (www.medisapiens.com)
ITGB6	RBMS1	UACC893	MediSapiens (www.medisapiens.com)
RARA	KIAA0195	UACC893	MediSapiens (www.medisapiens.com)
MED1	IKZF3	UACC893	MediSapiens (www.medisapiens.com)
FBXL20	CRKRS	UACC893	PMC3431177
GRB7	PPP1R1B	UACC893	PMC3431177/MediSapiens (www.medisapiens.com)
ESR1	CCDC170	ZR751	SOAPfuse
USP32	CCDC49	ZR7530	PMC3233654
DDX5	DEPDC6	ZR7530	PMC3233654
PLEC1	ENPP2	ZR7530	PMC3233654

BCAS3	HOXB9	ZR7530	PMC3233654
TAOK1	PCGF2	ZR7530	PMC3233654
APPBP2	PHF20L1	ZR7530	PMID: 25485619
COL14A1	SKAP1	ZR7530	PMID: 25485619
TRPS1	LASP1	ZR7530	PMID: 25485619
ERBB2	BCAS3	ZR7530	PMID: 25485619
ZMYM4	OPRD1	ZR7530	PMID: 25485619
TIMM23	ARHGAP32	ZR7530	PMID: 25485619
ATAD2	FAM178B	JIMT-1	PMID: 25485619
EXOSC1	CRTAC1	SUM149PT	PMC3233654

Appendix Table 7. Characteristics of breast cancer cell lines used for *in-silico* analysis

Cell lines	ER	HER2	ER/HER2
BT474	Pos	Pos	ER+/HER2+
MDAMB330	Pos	Pos	ER+/HER2+
MDAMB361	Pos	Pos	ER+/HER2+
UACC812	Pos	Pos	ER+/HER2+
ZR7530	Pos	Pos	ER+/HER2+
SUM52PE	Pos	Pos	ER+/HER2+
BT483	Pos	Neg	ER+/HER2-
CAMA1	Pos	Neg	ER+/HER2-
EFM19	Pos	Neg	ER+/HER2-
HCC1428	Pos	Neg	ER+/HER2-
MCF7	Pos	Neg	ER+/HER2-
MDAMB134	Pos	Neg	ER+/HER2-
MDAMB175V-II	Pos	Neg	ER+/HER2-
T47D	Pos	Neg	ER+/HER2-
HCC1008	Neg	Pos	ER-/HER2+
HCC1569	Neg	Pos	ER-/HER2+
HCC1954	Neg	Pos	ER-/HER2+
HCC202	Neg	Pos	ER-/HER2+

HCC2218	Neg	Pos	ER-/HER2+
JIMT-1	Neg	Pos	ER-/HER2+
KPL4	Neg	Pos	ER-/HER2+
MDAMB453	Neg	Pos	ER-/HER2+
SKBR3	Neg	Pos	ER-/HER2+
SUM225	Neg	Pos	ER-/HER2+
UACC893	Neg	Pos	ER-/HER2+
HCC1419	Neg	Pos	ER-/HER2+
BT20	Neg	Neg	ER-/HER2-
BT549	Neg	Neg	ER-/HER2-
HCC1143	Neg	Neg	ER-/HER2-
HCC1187	Neg	Neg	ER-/HER2-
HCC1395	Neg	Neg	ER-/HER2-
HCC1599	Neg	Neg	ER-/HER2-
HCC1806	Neg	Neg	ER-/HER2-
HCC1937	Neg	Neg	ER-/HER2-
HCC2157	Neg	Neg	ER-/HER2-
HCC3153	Neg	Neg	ER-/HER2-
HCC38	Neg	Neg	ER-/HER2-
MDAMB157	Neg	Neg	ER-/HER2-
MDAMB231	Neg	Neg	ER-/HER2-
MDAMB436	Neg	Neg	ER-/HER2-
MDAMB468	Neg	Neg	ER-/HER2-
SUM149PT	Neg	Neg	ER-/HER2-
ZR751	Neg	Neg	ER-/HER2-
HCC1493	NA	NA	NA
HCC2911	NA	NA	NA

3. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in cohort1, cohort 2, TCGA and METABRIC

Appendix Table 8. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in cohort 1.

Characteristic	n = 141	VMP1 mRNA level median (25th, 75th)	p-value
Age			
≥ 50	85	-0.23 (-0.59, 0.28)	0.67
< 50	56	-0.28 (-0.78, 0.57)	
Estrogen receptor			
positive	90	-0.20 (-0.59, 0.32)	0.41
negative	47	-0.39 (-1.03, 0.50)	
unknown	4		
Progesterone receptor			
positive	70	-0.19 (-0.57, 0.43)	0.42
negative	64	-0.31 (-0.97, 0.37)	
unknown	7		
HER2 status			
positive	23	0.58 (0.02, 1.26)	7x10 ⁻⁴ *
negative	117	-0.36 (-0.75, 0.18)	
unknown	1		
Receptors ER and HER2			
ER- and HER2-	32	-0.56 (-1.14, -0.10)	6.4x10 ⁻⁶ *
ER- and HER2+	14	0.51 (-0.13, 1.20)	
ER+ and HER2-	82	-0.24 (-0.59, 0.23)	
ER+ and HER2+	8	0.48 (0.29, 1.25)	
unknown	5		
Tumor size (mm)			
> 20	97	-0.31 (-0.69, 0.40)	0.9
≤ 20	44	-0.14 (-0.62, 0.31)	
Histological type			
IDC(a)	121	-0.14 (-0.65, -0.43)	0.24
ILC(b)	12	-0.35 (-0.56, -0.13)	
other	8	-0.37 (-0.74, -0.10)	
Nodal status			
positive	72	-0.20 (-0.55, 0.57)	0.16
negative	55	-0.13 (-0.88, 0.38)	

unknown	14		
Ki 67			
High	41	-0.46 (-0.85, 0.28)	0.15
Low	97	-0.17 (-0.53, 0.51)	
Unknown	3		
Histological grade			
1	12	-0.49 (-0.67, 0.26)	0.08
2	80	-0.15 (-0.52, 0.25)	
3	48	-0.25 (-0.80, 0.78)	
unknown	1		
Metastasis			
Positive	59	-0.14 (-0.52, 0.60)	0.03*
Negative	81	-0.36 (-0.75, 0.25)	
unknown	1		
Intrinsic subtype			
Basal	24	-0.75 (-1.22, -0.14)	5x10 ⁻⁶ *
ERBB2	14	0.82 (0.37, 1.35)	
Luminal A	43	-0.24 (-0.51, 0.17)	
Luminal B	30	0.03 (-0.56, 0.49)	
Normal-like	12	-0.47 (-0.87, -0.17)	
unknown	18		
Familial status			
BRCA2	27	-0.43 (-0.89, 0.39)	0.31
Non-BRCA2	114	-0.20 (-0.64, 0.38)	

The table shows the median and the 25th and 75th percentiles. One tumor was BRCA1 positive, and it was not used in the familial status calculations. The p-value is calculated with log2 transformed data using a t-test or ANOVA. *Significant difference $p < 0.05$.

Appendix Table 9. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in cohort 2

Characteristic	n=277	VMP1 mRNA level, median (25th, 75th)	p-value
Age			
≥ 50	213	-0.10 (-0.66, 0.42)	0.7
< 50	64	-0.32 (-0.92, 0.68)	
Estrogen receptor			
positive	194	-0.15 (-0.67, 0.45)	0.9
negative	77	-0.08 (-0.92, 0.42)	
unknown	6		
Progesterone receptor			
positive	176	-0.21 (-0.70, 0.41)	0.4
negative	94	-0.08 (-0.80, 0.56)	
unknown	7		
HER2 status			
positive	55	0.27 (-0.39, 0.98)	0.004*
negative	217	-0.17 (-0.80, 0.39)	
unknown	5		
Receptors ER and ERBB2			
ER neg and HER2 neg	38	-0.46 (-1.19, 0.10)	0.001*
ER neg and HER2 pos	23	0.27 (0.39, 1.02)	
ER pos and HER2 neg	133	-0.36 (-0.84, 0.30)	
ER pos and HER2 pos	18	0.08 (-0.51, 0.93)	
Tumor size (mm)			
> 20	194	-0.08 (-0.64, 0.50)	0.05
≤ 20	82	-0.26 (-0.81, 0.32)	
unknown	1		
Histological type			
IDC	231	-0.09 (-0.76, 0.51)	0.3
ILC	30	-0.41 (-0.72, 0.06)	
other	16	-0.11 (-0.64, 0.33)	
Nodal status			
positive	146	-0.22 (-0.74, 0.47)	0.6
negative	101	-0.07 (-0.77, 0.50)	
unknown	30		
Histological grade			
1	31	-0.35 (-0.82, 0.08)	0.04*
2	124	-0.11 (-0.64, 0.37)	
3	107	-0.07 (-0.79, 0.83)	
unknown	15		
Metastasis			
positive	65	0.00 (-0.76, 0.5)	0.7
negative	210	-0.12 (-0.65, 0.45)	
unknown	2		

The table shows the median and the 25th and 75th percentiles. The p-value was calculated with log2 normalized data using a t-test or ANOVA. *Significant difference $p < 0.05$.

Appendix Table 10. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in TCGA

Characteristic	n= 421	VMP1 mRNA level, median (25th, 75th)	p-value
Age			
≥ 50	125	0.16(-0.49, 1.18)	0.6
< 50	296	0.23 (-0.86, 1.29)	
Estrogen receptor			
positive	323	0.39 (-0.41, 1.34)	7x10 ⁻⁶ *
negative	92	-0.43 (-0.99, 0.28)	
unknown	6		
Progesterone receptor			0.008*
positive	273	0.42 (-0.33, 1.33)	
negative	141	-0.31 (-0.87, 0.55)	
unknown	7		
HER2 status			0.003*
positive	73	0.61 (-0.30, 1.85)	
negative	215	0.14 (-0.57, 0.96)	
unknown	133		
Histological type			0.8
IDC	360	0.22 (-0.50, 1.26)	
ILC	34	0.45 (-0.51, 1.22)	
unknown	27		
Nodal status			0.01*
positive	206	0.07 (-0.69, 1.08)	
negative	206	0.35 (-0.49, 1.32)	
unknown	9		
Subtype			2x10 ⁻¹² *
Basal	77	-0.58 (-1.17, -0.22)	
ERBB2	40	0.58 (-0.30, 2.02)	
Luminal A	210	0.38 (-0.47, 1.23)	
Luminal B	89	0.68 (-0.30, 1.92)	
Normal-like	4	0.36 (-0.29, 0.68)	
unknown	1		

The table shows the median and the 25th and 75th percentiles. The p-value was calculated with normalized Z-scores (A_23_P129935) using a t-test or ANOVA.

*Significant difference $p < 0.05$.

Appendix Table 11. Correlation of VMP1 mRNA with clinicopathological characteristics of breast tumors in METABRIC

Characteristic	n=1904	VMP1 mRNA level, median (25th, 75th)	p-value
Age			
≥ 50	1493	0.18 (-0.52, 1.00)	0.7
< 50	411	0.07 (-0.51, 0.97)	
Estrogen receptor			
positive	1459	0.24 (-0.48, 1.04)	0.01*
negative	445	-0.04 (-0.69, 0.87)	
Progesterone receptor			
positive	1009	0.16 (-0.51, 0.97)	0.5
negative	895	0.15 (-0.54, 1.05)	
HER2 status			
positive	236	0.97 (0.14, 2.20)	<2x10-16*
negative	1668	0.07 (-0.61, 0.85)	
Histological type			
IDC	1502	0.18 (-0.51, 1.05)	0.02*
ILC	141	-0.04 (-0.76, 0.73)	
Other	261	0.15 (-0.5, 0.99)	
Subtype			
Basal	199	-0.38 (-1.02, 0.37)	<2x10-16*
ERBB2	220	0.66 (-0.09, 1.65)	
Luminal A	679	0.09 (-0.53, 0.78)	
Luminal B	461	0.47 (-0.43, 1.26)	
Normal-like	140	0.22 (-0.62, 1.02)	
claudin-low	199	0.09 (-0.52, 0.91)	
unknown	6		

The table shows the median and the 25th and 75th percentiles. The p-value was calculated with normalized Z-scores (Illumina Human v3 microarray) using a t-test or ANOVA. *Significant difference $p < 0.05$.

4. Correlation of hsa-miR-21-3p with clinicopathological characteristics of breast tumors in cohort 1, cohort 2 and TCGA

Appendix Table 12. Correlation of hsa-miR-21-3p with clinicopathological characteristics of breast tumors in cohort 1

Characteristic	n=139	miR21-3p mRNA median (25 and 75%)	p-value
Age			0.263
< 50	53	0.114 (-0.420, 1.277)	
≥ 50	86	-0.157 (-1.007, 0.815)	
Estrogen receptor			0.483
Negative	46	-0.261 (-0.988, 0.818)	
Positive	89	0.116 (-0.677, 1.104)	
Unknown	4		
Progesterone receptor			0.361
Negative	63	-0.247 (-0.910, 0.801)	
Positive	69	0.1271 (-0.677, 1.201)	
Unknown	7		
HER2 status			0.286
Negative	116	-0.041 (-0.921, 1.025)	
Positive	22	0.239 (-0.330, 1.172)	
Unknown	1		
Tumor size			0.340
≤ 20	45	0.269 (-0.517, 1.018)	
> 20	94	-0.157 (-0.974, 0.997)	
Unknown	0		
1			0.473
2	12	0.130 (-0.439, 1.327)	
3	79	0.042 (-1.021, 1.032)	
Unknown	47	-0.063 (-0.625, 0.774)	
	1		
Ki67			0.247
1	94	0.060 (-0.724, 1.071)	
2	28	-0.354 (-1.166, 0.395)	
3	15	-0.247 (-0.452, 0.750)	
Unknown	2		
Nodes			0.573
Negative	55	-0.063 (-0.862, 0.936)	
Positive	67	0.021 (-0.733, 1.186)	
Unknown	17		
Histology subtype			0.305

Ductal	120	-0.036 (-0.890, 0.905)	
Ductal_lobular	1	0.854	
Ductal_mucinous	1	0.713	
Lobular	11	0.429 (-0.069, 1.927)	
Medullary	4	-0.749 (-1.110, -0.061)	
Metaplastic	0	NA	
Metastasis_adenocarcinoma_breast_mucino us	1	(-0.997)	
Mucinous	1	(-1.136)	
Subtype Hu et al.			0.462
Basal	24	-0.361 (-0.975, 0.430)	
ERBB2	13	0.114 (-0.275, 0.817)	
LumA	42	0.159 (-0.902, 0.969)	
LumB	29	-0.059 (-1.161, 1.104)	
Normal	11	0.384 (-0.443, 2.594)	
unclassified	12	-0.053 (-0.694, 1.324)	
Unknown	6		
BRCA family status			0.659
BRCA2	27	-0.059 (-0.931, 1.262)	
BRCAX	67	0.116 (-0.391, 0.856)	
BRCAX-like	6	0.057 (-0.680, 0.285)	
Sporadic	37	-0.483 (-1.323, 1.080)	
Metastasis			0.021
M0	62	-0.416 (-1.041, 0.691)	
M1	76	0.307 (-0.358, 1.306)	
Unknown	1		

Appendix Table 13. Correlation of hsa-miR-21-3p with clinicopathological characteristics of breast tumors in cohort 2

Characteristic	n=281	miR21-3p mRNA median (25 and 75%)	p-value
Age			0.965
< 50	64	0.014 (-0.659, 0.584)	
≥ 50	217	-0.009 (-0.554, 0.636)	
Estrogen receptor			0.036
Negative	77	0.191 (-0.448, 0.920)	
Positive	197	-0.037 (-0.730, 0.569)	
Unknown	7		
Progesterone receptor			0.024
Negative	94	0.194 (-0.456, 0.844)	
Positive	179	-0.092 (-0.781, 0.585)	
Unknown	8		
ERBB2 status			0.019
Negative	175	-0.080 (-0.786, 0.538)	
Positive	46	0.469 (-0.329, 1.317)	
Unknown	60		
HER2 status			0.003
Negative	217	-0.022 (-0.759, 0.551)	
Positive	51	0.398 (-0.341, 1.195)	
NA	13		
Tumor size			0.078
≤ 20	83	-0.104 (-0.782, 0.546)	
> 20	196	0.016 (-0.528, 0.672)	
Unknown	2		
Histologic Grade			0.026
1	31	-0.080 (-0.841, 0.365)	
2	128	-0.083 (-0.737, 0.647)	
3	111	0.208 (-0.526, 0.744)	
Unknown	11		
Unknown			
Nodes			0.766
Negative	104	0.014 (-0.692, 0.749)	
Positive	146	0.004 (-0.551, 0.612)	
Unknown	31		
Histology subtype			0.457
Ductal	234	0.014 (-0.544, 0.712)	
Ductal_lobular	6	0.252 (-0.153, 0.553)	
Ductal_mixed	1	-0.759	
Lobular	29	-0.209 (-1.136, 0.420)	
Lobular_mixed	1	1.26	
Medullary	1	-0.042	
Metaplastic	2	0.788 (0.596, 0.979)	
Mucinous	5	0.022 (-1.454, 0.176)	
Sarcoma	1	(-0.328)	
Tubular	1	(-1.254)	

Appendix Table 14. Correlation of hsa-miR-21-3p with clinicopathological characteristics of breast tumors in TCGA

Characteristic	n=256	miR21-3p mRNA median (25 and 75%)	p-value
Age			0.142
< 50	61	-0.038 (-0.526, 0.754)	
≥ 50	195	-0.067 (-0.776, 0.565)	
Estrogen receptor			0.767
Negative	57	-0.014 (-0.769, 0.583)	
Positive	196	-0.066 (-0.688, 0.563)	
Unknown	3		
Progesterone receptor			0.578
Negative	88	-0.144 (-0.931, 0.550)	
Positive	164	-0.045 (-0.642, 0.574)	
Unknown	4		
HER2 status			0.001
Negative	217	-0.138 (-0.770, 0.470)	
Positive	35	0.510 (-0.207, 1.253)	
Unknown	4		
Tumor stage			0.152
T1	51	-0.318 (-0.806, 0.203)	
T2	154	0.069 (-0.624, 0.685)	
T3	34	-0.470 (-0.804, 0.335)	
T4	14	0.238 (-0.290, 0.785)	
TX	3	0.323 (-0.224, 0.331)	
Nodes			0.683
Negative	126	-0.055 (-0.618, 0.658)	
Positive	130	-0.134 (-0.801, 0.506)	
Nodal status			0.361
N0	126	-0.055 (-0.618, 0.658)	
N1	73	-0.131 (-0.943, 0.308)	
N2	36	-0.200 (-0.72, 0.441)	
N3	21	0.129 (-0.326, 0.833)	
Metastasis			0.314
M0	243	-0.065 (-0.710, 0.565)	
M1	9	0.049 (-0.181, 0.837)	
Unknown	4		
Subtype PAM50*			2.98E-05
Basal-like	43	-0.131 (-0.933, 0.439)	
HER2-enriched	29	0.355 (-0.312, 1.269)	
Luminal A	106	-0.231 (-0.951, 0.176)	

Luminal B	73	0.261 (-0.419, 0.821)	
Normal-like	5	0.323 (-0.404, 0.499)	
Receptor status			1.27E-05
ER+HER2+	25	0.808 (0.209, 1.960)	
ER+HER2-	168	-0.178 (-0.768, 0.428)	
ER-HER2+	9	-0.183 (-0.768, 0.355)	
ER-HER2-	47	-0.014 (-0.772, 0.646)	

Correlation of hsa-miR-21-5p with clinicopathological characteristics of breast tumors in cohort1, cohort 2, TCGA and METABRIC/EGA

Appendix Table 15. Correlation of has-miR-21-5p with clinicopathological characteristics of breast tumors in cohort 1

Characteristic	n=140	miR21-5p mRNA median (25 and 75%)	p-value
Age			0.534
< 50	53	-0.021 (-0.391, 0.437)	
≥ 50	87	0.008 (-0.576, 0.416)	
Estrogen receptor			0.723
Negative	47	0.072 (-0.666, 0.650)	
Positive	89	-0.036 (-0.416, 0.335)	
Unknown	4		
Progesterone receptor			0.695
Negative	64	0.061 (-0.520, 0.469)	
Positive	69	-0.036 (-0.478, 0.411)	
Unknown	7		
HER2 status			0.002
Negative	116	-0.046 (-0.605, 342)	
Positive	23	0.395 (-0.008, 0.809)	
Unknown	1		
Tumor size			0.363
≤ 20	45	-0.097 (-0.666, 0.361)	
> 20	95	0.016 (-0.405, 0.497)	
Unknown	0		
Histologic Grade			0.977
1	12	-0.038 (-0.157, 0.397)	
2	79	0.011 (-0.447, 0.311)	
3	48	0.103 (-0.656, 0.597)	
Unknown	1		
Ki67			0.610
1	95	-0.008 (-0.414, 0.333)	
2	27	0.293 (-0.501, 0.585)	
3	15	-0.106 (-0.657, 0.728)	
Unknown	3		
Nodes			0.432
Negative	55	-0.046 (-0.595, 0.371)	
Positive	70	0.109 (-0.393, 0.507)	
Unknown	15		

Histology subtype			0.744
Ductal	121	-0.034 (-0.548, 0.437)	
Ductal_lobular	1	0.016	
Ductal_mucinous	1	0.414	
Lobular	11	0.145 (-1.239, 0.302)	
Medullary	4	0.529 (0.148, 0.882)	
Metaplastic	0	NA	
Metastasis_adenocarcinoma_breast_mucinous	1	0.164	
Mucinous	1	(-0.157)	
Subtype Hu et al.			0.261
Basal	24	-0.129 (-0.829, 0.625)	
ERBB2	14	0.386 (0.105, 0.678)	
LumA	42	-0.046 (-0.454, 0.330)	
LumB	29	-0.036 (-0.351, 0.331)	
Normal	11	-0.021 (-0.595, 0.130)	
unclassified	12	0.228 (-0.505, 0.479)	
Unknown	8		
BRCA family status			0.264
BRCA1	2	0.075 (-0.361, 0.511)	
BRCA2	27	-0.236 (-0.585, -0.029)	
BRCAX	67	0.051 (-0.592, 0.403)	
BRCAX-like	6	0.265 (0.085, 0.786)	
Sporadic	38	-0.175 (-0.454, 0.593)	
Metastasis			0.394
M0	62	0.158 (-0.537, 0.471)	
M1	77	-0.070 (-0.498, 0.411)	
Unknown	1		

Appendix Table 16. Correlation of hsa-miR-21-5p with clinicopathological characteristics of breast tumors in cohort 2

Characteristic	n=282	miR21-5p mRNA median (25 and 75%)	p-value
Age			0.769
< 50	64	-0.011 (-0.626, 0.498)	
≥ 50	218	0.023 (-0.613, 0.541)	
Estrogen receptor			0.652
Negative	77	0.166 (-0.579, 0.506)	
Positive	198	-0.021 (-0.621, 0.533)	
Unknown	7		
Progesterone receptor			0.772
Negative	94	0.146 (-0.622, 0.546)	
Positive	180	-0.029 (-0.610, 0.516)	
Unknown	8		
ERBB2 status			0.799
Negative	176	-0.033 (-0.626, 0.530)	
Positive	46	0.236 (-0.384, 0.643)	
Unknown	60		
HER2 status			0.209
Negative	218	-0.029 (-0.627, 0.475)	
Positive	51	0.232 (-0.418, 0.606)	
NA	13		
Tumor size			0.416
≤ 20	84	-0.045 (-0.704, 0.493)	
> 20	196	0.023 (-0.499, 0.548)	
Unknown	2		
Histologic Grade			0.609
1	31	-0.052 (-0.642, 0.500)	
2	129	-0.007 (-0.546, 0.476)	
3	111	0.123 (-0.634, 0.567)	
Unknown	11		
Nodes			0.693
Negative	104	-0.000 (-0.587, 0.559)	
Positive	147	0.033 (-0.632, 0.518)	
Unknown	31		
Histology subtype			0.263
Ductal	234	0.075 (-0.600, 0.546)	
Ductal_lobular	6	-0.207 (-0.845, 0.319)	
Ductal_mixed	1	-1.327	
Lobular	30	-0.188 (-0.471, 0.288)	
Lobular_mixed	1	1.340	
Medullary	1	-0.420	
Metaplastic	2	0.040 (-0.412, 0.492)	
Mucinous	5	-0.244 (-0.916, 0.791)	
Sarcoma	1	(-2.615)	
Tubular	1	(-0.678)	

Appendix Table 17. Correlation of hsa-miR-21-5p with clinicopathological characteristics of breast tumors in TCGA

Characteristic	n=283	miR21-5p mRNA median (25 and 75%)	p-value
Age			0.008
< 50	66	0.130 (-0.481, 0.749)	
≥ 50	217	-0.069 (-0.828, 0.468)	
Estrogen receptor			0.665
Negative	61	0.037 (-0.810, 0.632)	
Positive	218	-0.020 (-0.705, 0.519)	
Unknown	4		
Progesterone receptor			0.267
Negative	99	-0.050 (-0.886, 0.486)	
Positive	179	0.010 (-0.658, 0.560)	
Unknown	5		
HER2 status			0.094
Negative	236	-0.059 (-0.797, 0.515)	
Positive	36	0.280 (-0.506, 0.692)	
Unknown	11		
Tumor stage			0.429
T1	59	0.085 (-0.694, 0.580)	
T2	169	-0.025 (-0.704, 0.521)	
T3	35	-0.322 (-1.063, 0.348)	
T4	17	0.100 (-0.296, 0.627)	
TX	3	0.145 (-0.015, 0.440)	
Nodes			0.618
Negative	140	0.036 (-0.730, 0.583)	
Positive	143	-0.025 (-0.720, 0.504)	
Nodal status			0.516
N0	140	0.036 (-0.730, 0.583)	
N1	84	-0.014 (-0.777, 0.559)	
N2	37	0.000 (-0.501, 0.535)	
N3	22	-0.314 (-0.851, 0.285)	
Metastasis			0.418
M0	269	0.000 (-0.724, 0.546)	
M1	10	-0.102 (-0.409, 0.650)	
Unknown	4		
Subtype PAM50*			0.460
Basal-like	48	-0.293 (-1.218, 0.533)	
HER2-enriched	31	0.010 (-0.610, 0.434)	

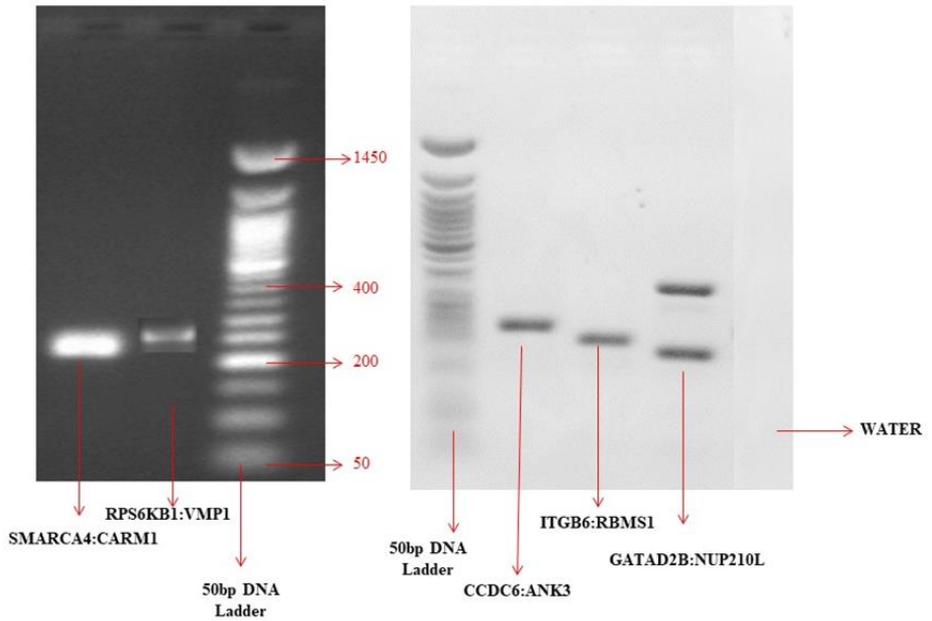
Luminal A	116	0.080 (-0.566, 0.487)
Luminal B	77	-0.025 (-0.942, 0.674)
Normal-like	5	-0.340 (-0.757, 0.312)
Unknown	6	

Appendix Table 18. Correlation of hsa-miR-21-5p with clinicopathological characteristics of breast tumors in METABRIC/EGA

Characteristic	n=1220	miR21-5p mRNA median (25 and 75%)	p-value
Age			
< 50	279	-0.028 (-0.512, 0.391)	0.715
≥ 50	941	0.008 (-0.526, 0.415)	
Estrogen receptor			0.057
Negative	281	0.096 (-0.433, 0.530)	
Positive	939	-0.017 (-0.539, 0.380)	
Progesterone receptor			0.003
Negative	581	0.043 (-0.452, 0.479)	
Positive	639	-0.038 (-0.580, 0.363)	
Unknown	3		
HER2 status			1.01E-14
Negative	1067	-0.063 (-0.570, 0.345)	
Positive	153	0.415 (-0.420, 0.786)	
Unknown	0		
Tumor stage			0.382
1	364	0.025 (-0.473, 0.401)	
2	588	-0.029 (-0.527, 0.362)	
3	98	0.067 (-0.534, 0.464)	
4	10	0.354 (-0.057, 0.642)	
Unknown	158		
Tumor size			0.658
≤ 20	527	0.048 (-0.431, 0.418)	
> 20	680	-0.036 (-0.541, 0.400)	
Unknown	13		
Histologic Grade			0.210
1	106	0.085 (-0.268, 0.345)	
2	494	-0.089 (-0.589, 0.373)	
3	620	0.048 (-0.468, 0.461)	
Unknown	1220		
Cellularity			0.050
Low	137	-0.148 (-0.624, 0.287)	
Moderate	447	0.006 (-0.469, 0.391)	

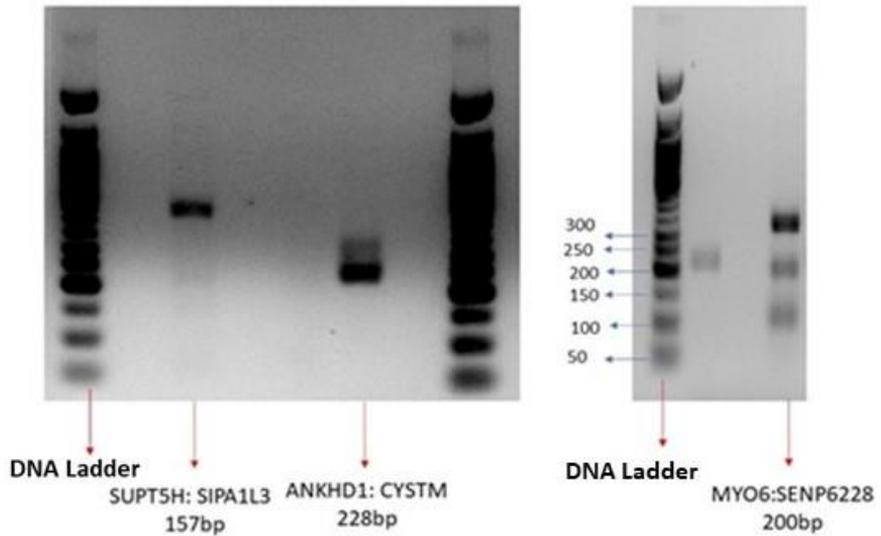
High	586	0.064 (-0.484, 0.457)	
Unknown	50		
Nodes			0.380
Negative	623	0.001 (-0.529, 0.405)	
Positive	551	-0.008 (-0.505, 0.415)	
Unknown	46		
Nodal status			0.484
Negative	623	0.001 (-0.529, 0.405)	
Positive	358	-0.037 (-0.471, 0.368)	
Unknown	127	-0.017 (-0.609, 0.431)	
	66	-0.008 (-0.505, 0.415)	
	46		
Histology subtype			5.75E-06
Ductal/NST	922	0.041(-0.441, 0.452)	
Loublar	96	-0.438 (-1.014, 0.265)	
Medullary	15	0.149 (-0.520, 0.373)	
Metaplastic	0	0	
Mixed	139	-0.047 (-0.500, 0.316)	
Mucinous	12	-0.763 (-1.247, -0.229)	
Other	7	0.610 (-1.427, 0.781)	
Tubular/cribriform	15	-0.065 (-0.273, 0.223)	
Unknown	14		
Subtype PAM50			3.31E-05
Basal	118	-0.061 (-0.814, 0.342)	
Claudin-low	148	-0.038 (-0.551, 0.397)	
Her2	109	0.228 (-0.319, 0.755)	
LumA	385	0.000 (-0.446, 0.324)	
LumB	294	0.017 (-0.551, 0.469)	
NC	4	-0.124 (-0.178, 0.042)	
Normal	102	-0.090 (-0.870, 0.388)	
Unknown	60		
3-Gene Classifier Subtype			2.28E-09
ER-/HER2-	196	-0.009 (-0.620, 0.462)	
ER+/HER2- High Prolif	394	-0.029 (-0.542, 0.385)	
ER+/HER2- Low Prolif	411	-0.018 (-0.513, 0.302)	
HER2+	127	0.388 (-0.113, 0.804)	
Unknown	92		

5. Appendix figures



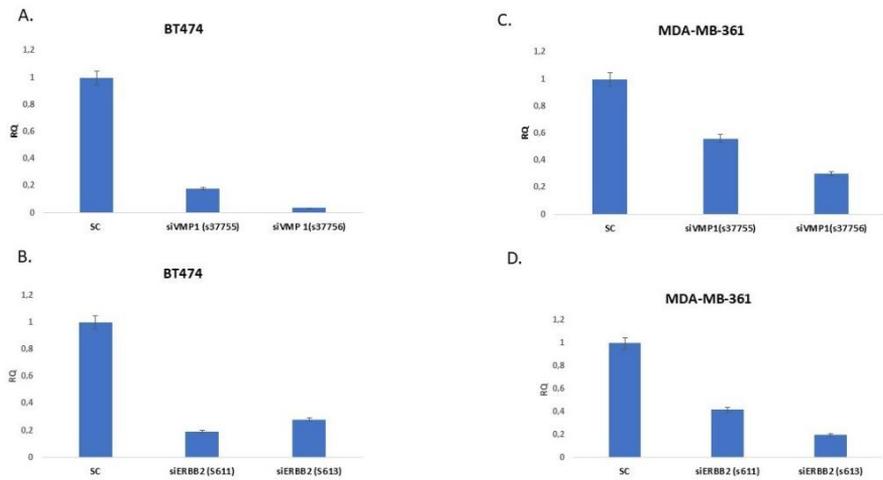
Appendix Figure 1. Validated fusion genes with RT-PCR

The fusion genes were amplified with PCR. The PCR products were run on 2% agarose gel at 100V for 40 minutes. A DNA ladder is shown on the side for reference. The primers were designed to span the junction of the fusion gene. All the PCR products were of the expected size.



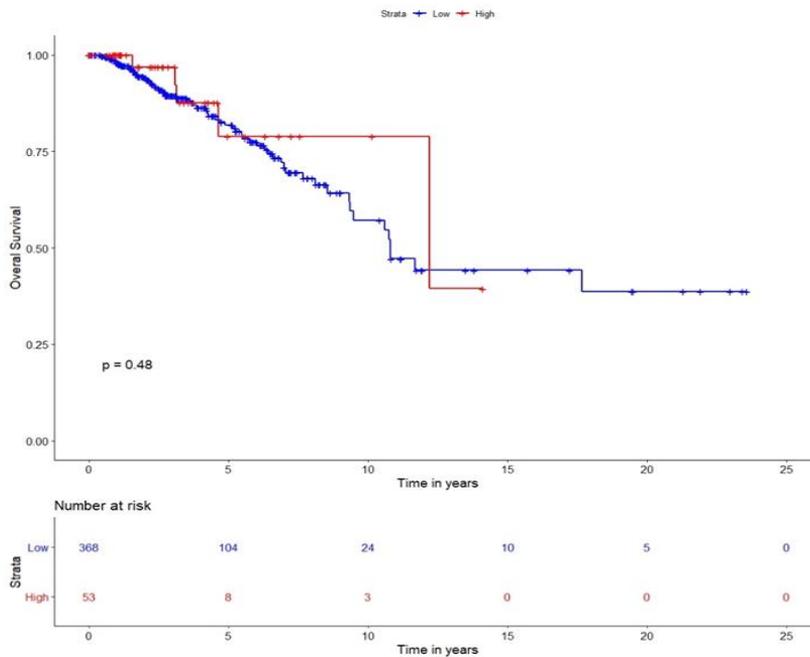
Appendix figure 2. Validated fusion genes with RT-PCR

The fusion genes were amplified with PCR. The PCR products were run on 2% agarose gel at 100V for 40 minutes. A DNA ladder is shown on the side for reference. The primers were designed to span the junction of the fusion gene. All the PCR products were of the expected size. The empty lanes were water with no template used as negative control.



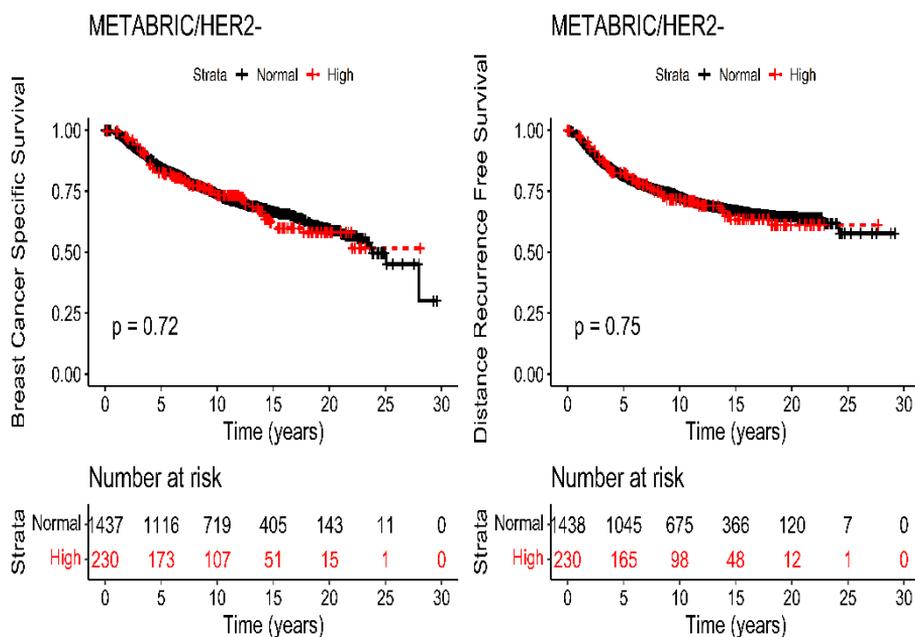
Appendix figure 3: q-PCR confirmations of silencing VMP1 and ERBB2 genes

A, B,C,D: in A) and B) BT474 and C) and D) MDA-MB-361 cell lines by using two different siRNA assays for VMP1 (s37755) and (s37756) and (s611) and (s613) for ERBB2 gene. Seventy-two hours after adding siRNAs cells were collected for protein and mRNA extraction. Both VMP1 and ERBB2 mRNA quantifications showed in each figure. Both genes quantifications were normalized based on SCRAMBLE. A and B representing data from BT474 cell line and C and D from MDA-MB-361 cell line.



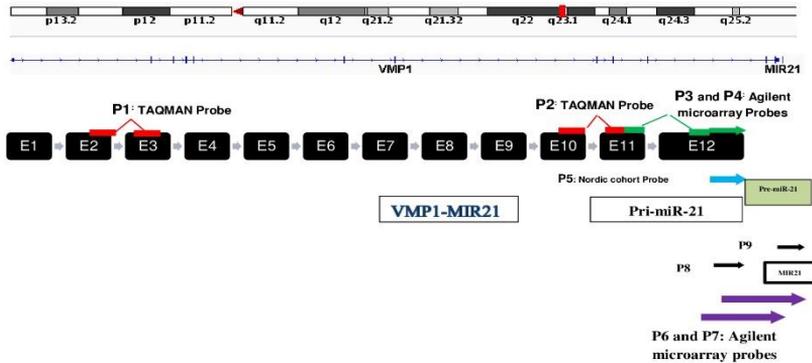
Appendix figure 4. High RPS6KB1 was not marker of shorter overall survival

Based on RPS6KB1 mRNA, patient tumors divided to 2 groups: tumors with high RPS6KB1 (high \geq mean + 1 SD) and normal RPS6KB1 (normal $<$ mean + 1 SD). Overall survival (OS) of patients in TCGA was examined in respect to quantity of RPS6KB1 mRNA. In Nordic cohort there was not data for RPS6KB1 mRNA.



Appendix figure 5 High expression of VMP1 was not associated with shorter BCSS and DRFS within METABRIC/ HER2 negative tumors

In HER2 negative breast cancer patients from METABRIC, patients categorized to high VMP1 (high \geq mean + 1 SD) and normal VMP1 (normal < mean + 1 SD) and analyzed with respect to breast cancer specific survival (BCSS) and distant recurrence free survival (DRFS). The log rank p-values are shown in the figure and the numbers of patients at risk are shown in the table below the graphs



Appendix figure 6. Location of probes within VMP1 and MIR21. The schematic at top shows chromosome 17 with a red square that indicates the location of the *VMP1* and *MIR21* genes. Below is a stick diagram of the exons within *VMP1* as well as the mature sequence of MIR-21. In the box diagram are shown the exons and the probes used to measure *VMP1* mRNA levels. Probe P1 denotes the Taqman probe that spans exons 2 and 3, which was used in cohort 1 to check if *MIR21* influenced the measurement of *VMP1* levels. P2 denotes the Taqman probe that spans exons 10 and 11, which was used in cohorts 1 and 2. P3 and P4 denote the microarray probes from the cancer genome atlas (TCGA) and the METABRIC cohorts. P5 shows the location of the microarray probe in the Nordic cohort. P6 and P7 denote the Agilent microarray probes from TCGA and European-Phenome Genome Archive (EGA) cohort1. P8 and P9 denote miRCURY LNA miRNA detection probes in cohort 1 and cohort 2. The box underneath labeled *MIR21* shows the position of the *MIR21* gene. Note that the mature sequence of *MIR21* is telomeric to the *VMP1* gene.