Potential oncogenes within the 8p12-p11 amplicon

Identification and functional testing in breast cancer cell lines

Edda Olgudóttir

Thesis for the degree of Master of Science
University of Iceland
Faculty of Medicine
School of Health Science

HÁSKÓLI ÍSLANDS

Hugsanleg æxlisgen á 8p12-p11 mögnunarsvæðinu

Æxlisgen tilgreind og prófuð í brjóstakrabbameinsfrumulínum

Edda Olgudóttir

Ritgerð til meistaragráðu í Líf- og læknavísindum

Leiðbeinendur: Inga Reynisdóttir og Rósa Björk Barkardóttir

Meistaranámsnefnd: Inga Reynisdóttir, Rósa Björk Barkardóttir

og Guðmundur Hrafn Guðmundsson

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Febrúar 2012

Potential oncogenes within the 8p12-p11 amplicon

Identification and functional testing in breast cancer cell lines

Edda Olgudóttir

Thesis for the degree of Master of Science

Supervisor: Inga Reynisdóttir and Rósa Björk Barkardóttir

Masters committee: Inga Reynisdóttir, Rósa Björk Barkardóttir and Guðmundur Hrafn Guðmundsson

Faculty of Medicine

School of Health Sciences

February 2012

Prentun: Háskólaprent

Staður, Ísland 2012

Ágrip

Orsakir brjóstakrabbameins eru margþættar og margar þeirra leiða til brenglana í erfðaefninu. Algengar erfðabrenglanir í brjóstakrabbameini eru tap eða magnanir á litningasvæðum. Svæði sem eru gjörn á að tapast innihalda svokölluð æxlisbæligen en svæði sem magnast innihalda æxlisgen, gen sem ýta undir æxlisvöxt. Æxlisgen sem eru staðsett á mögnunarsvæði eru oft yfirtjáð samhliða mögnun þeirra og hafa þar af leiðandi aukna virkni. Vegna þessa má líta svo á að æxlisgenin drífi mögnunina áfram og þau séu því markgen mögnunarinnar. Litningasvæðið 8p12-p11 finnst magnað í u.þ.b. 15-20% brjóstakrabbameina. Hingað til hefur einungis tekist að skilgreina markgen mögnunarinnar í einum undirflokk brjóstakrabbameina, luminal B undirflokk. Hins vegar er 8p12-p11 magnað í fleiri undirflokkum sem virðast ekki nýta sama kerfi og luminal B, hvað varðar þetta mögnunarsvæði. Í verkefni sem unnið var á Frumulíffræðideild við Rannsóknarstofu í Meinafræði við Landspítala Háskólasjúkrahús var notast við íslensk brjóstaæxlissýni til að þrengja mögnunarsvæðið og skilgreina hugsanleg markgen mögnunarinnar. Af 20 genum sem staðsett eru á svæðinu voru 11 genanna útilokuð sem markgen mögnunarinnar vegna skorts á aukinni tjáningu við mögnun. Sjö genanna sýndu aukna mRNA og prótíntjáningu við mögnun og voru því skilgreind sem hugsanleg æxlisgen. Tvö genanna, LETM2 og PPAPDC1B, voru ekki prófuð með tilliti til prótíntjáningar, þar sem mótefni voru ekki til. Í því verkefni sem lýst verður í þessari ritgerð var prótíntjáning LETM2 og PPAPDC1B prófuð í 17-36 brjóstakrabbameinssýnum. Hvorki LETM2 né PPAPDC1B sýndu aukna tjáningu samhliða mögnun. Sjö af tuttugu genum koma því til greina sem æxlisgen: ERLIN2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1 og DDHD2. Fjögur þessara gena, ERLIN2, PROSC, ASH2L og LSM1, voru prófuð frekar í því verkefni sem lýst er í þessari ritgerð. Til að athuga æxlismyndandi eiginleika genanna var notast við siRNA til að bæla tjáningu þeirra í brjóstakrabbameinsfrumulínum sem eru með mögnun og yfirtjáningu á 8p12-p11. Lifun frumnanna eftir bælingu var metin með MTT-prófi og talningu. Bæling á genunum hafði ekki áhrif á lifun brjóstakrabbameinsfrumulínanna. Að auki var leitast við að athuga samverkandi áhrif gena á mögnunarsvæðinu, með því að bæla tjáninguna á þremur genum samtímis. ASH2L, LSM1 og PROSC eru oft mögnuð og yfirtjáð samtímis í æxlum og voru öll þrjú genin því bæld í brjóstakrabbameinsfrumulínu, samtímis. Við þessar tilraunir var einnig notað siRNA og lifunin metin með talningu. Bæling á þremur genum samtímis hafði ekki áhrif á lifun brjóstakrabbameinsfrumulínunnar sem notuð var við þessa tilraun. Niðurstöður þessarar rannsóknar sýndu að fjögur þeirra sjö gena sem koma til greina sem hugsanleg æxlisgen á 8p12-p11 mögnunarsvæði, ERLIN2, PROSC, ASH2L og LSM1, hafa ekki áhrif á lifun brjóstakrabbameinsfrumulína.

Abstract

Many factors are known to contribute to breast cancer development. Many of them lead to genetic alterations. Common alterations in breast cancer are loss or amplifications of chromosomal regions. Chromosomal regions that tend to be lost contain tumor suppressor genes but regions that are commonly gained or amplified contain oncogenes, genes that promote tumor growth. Oncogenes that are located within amplified regions often increase their function through overexpression. Because of this, oncogenes within amplicons are also defined as target genes of the amplification. Chromosomal region 8p12-p11 is found gained or amplified in 15-20% of breast tumors. The target gene of this amplicon has been identified in one breast cancer subtype, the luminal B subtype. However this amplification is found in other subtypes as well. These subtypes do not seem to depend on the same mechanism as luminal B. Minimal amplified region and its candidate target genes had been identified in a previous project performed at the Cell Biology unit of the Department of Pathology at Landspitali-University Hospital in a set of breast tumors originating from Icelandic breast cancer patients. Of the 20 genes located within the region 11 genes were excluded as target genes of the amplification based on lack of increased expression. Seven genes showed increased mRNA and protein expression when amplified and were listed as putative targets of the amplification. Two genes, LETM2 and PPAPDC1B, were not tested due to lack of commercial antibodies at the time. The project described in this thesis aimed to test the protein expression of LETM2 and PPAPDC1B in 17 - 36 breast tumor samples. Neither LETM2 nor PPAPDC1B showed increased protein expression when amplified. LETM2 and PPAPDC1B were therefore excluded as target genes of the amplification. The seven remaining candidate target genes are: ERLIN2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1 and DDHD2. Four of these genes were tested further in the project described in this thesis. They are: ERLIN2, PROSC, ASH2L and LSM1. To estimate the oncogenic abilities of these genes, siRNA was used to downregulate the gene expression in breast cancer cell lines harboring amplification and overexpression of the 8p12-p11 region. Cell proliferation was evaluated after downregulation using MTT-assay and cell counting. Downregulation of the genes did not affect the cell proliferation. In addition, synergistic effects of genes co-amplified on the amplicon were tested. ASH2L, LSM1 and PROSC are often amplified and overexpressed simultaneously in tumors. Therefore these three genes were downregulated via siRNA in a breast cancer cell line, simultaneously, and proliferation estimated by counting. The downregulation of the three genes simultaneously did not affect the survival of the breast cancer cell line used in this experiment. The results show that the four of the seven candidate genes tested in this project do not affect cell proliferation in the breast cancer cell lines used for this study.

Acknowledgements

I would like to thank my supervisors Inga Reynisdóttir and Rósa Björk Barkardóttir for giving me the opportunity to work on this project and for their excellent guidance throughout the process of this project. I would also like to thank the third person of my MSc committee Guðmundur Hrafn Guðmundsson for contributing his time and resources to support this project.

To my coworkers at the Department of Cell biology, Aðalgeir Arason, Berglind Ósk Einarsdóttir, Eydís Þórunn Guðmundsdóttir, Guðrún Jóhannesdóttir, Haukur Gunnarsson, Óskar Örn Hálfdánarson and Þórgunnur Eyfjörð Pétursdóttir, I would like to express my gratitude for all the support and helpful discussions during our collaboration.

I would like to thank the team of Þórarinn Guðjónsson and Magnús Karl Magnússon at the Stem Cell Research Unit at Læknagarður, for letting me use their facilities, for their stimulating discussions and for materials provided by them, special thanks to Ari Jón Arason, Bylgja Hilmarsdóttir, Sævar Ingþórsson and Valgarður Sigurðson for their technical help. In addition I would like to thank the team of Pétur Henry Petersen at the Neurobiology at Læknagarður for providing me with western blot material and knowledge.

Last but not least I would like to express my deepest gratitude to my family and friends for their support and tolerance during this process. Very special thanks to Hörður Bjarnason and my daughter Ester Lind Gunnarsdóttir.

This project was supported by funding from The Icelandic Research Fund (Rannsóknarnámssjóður Rannís), The Research Fund of Landspítali-University Hospital (Vísindasjóður Landspítalans Háskólasjúkrahús) and The Icelandic association “Walking for Breast Cancer” (Göngum saman)

Table of contents

List of figures

List of tables

Abbreviations

ABL - c-ABL oncogene 1, non-receptor tyrosine kinase

aCGH - array Comparative genomic hybridization

AP2δ - Activating protein 2 δ

ASH2L - Ash2 (absent, small or homeotic) - like (Drosophila)

BAC - Bacterial artificial chromosome

BASE - BioArray Software Environment

BCR - Breakpoint cluster region

BFB - Breakage-fusion-bridge

BRCA1 - Breast cancer 1, early onset

BRCA2 Breast cancer 2, early onset

BRF1 - TFIIIB-related factor 1

BRF2 - TFIIIB-related factor 2

BRIP1 - BRCA1 interacting protein C-terminal helicase 1

CASP8 - Caspase 8, apoptosis-related cysteine peptidase

CASP10 - Caspase 10, apoptosis-related cysteine peptidase

CCDS - Consensus coding sequences

CCND1 - Cyclin D1

CHEK2 - Check point kinase 2

C8orf68 - Chromosome 8 open read frame 68

DCP1A - DCP1 decapping enzyme homolog A (S. cerevisiae)

DCP2 - DCP2 decapping enzyme homolog 2 (S. cerevisiae)

DDHD2 - DDHD domain containing 2

DGPP - Diacylglycerol pyrophosphate

DMs - Double minutes

DPP1 - Diacylglycerol pyrophosphate phosphatase

DPPL1 - Diacylglycerol pyrophosphate phosphatase like 1

DPY-30 - Dpy-30 homolog (C. elegans)

EGFR - Epidermal growth factor receptor

EMT - Epithelial to mesenchymal transition

ER - endoplasmic reticulum, in association with ERLIN 1 and 2 otherwise ER refers to the estrogen receptor

ER - Estrogen receptor

ERBB2 - v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)

ERBB3 - v-erb-b3 erythroblastic leukemia viral oncogene homolog 3 (avian)

ERLIN1 - ER-lipid raft associated 1 (Endoplasmic reticulum-lipid raft associated 1)

ERLIN2 - ER- lipid raft associated 2 (Endoplasmic reticulum-lipid raft associated 2)

EST - Expressed sequenced tag

FBS - Fetal Bovine Serum

FGFR1 - Fibroblast growth factor receptor 1

FGFR2 - Fibroblast growth factor receptor 2

FISH - Fluorescent in situ hybridization

GAB2 - GRB2 associated binding protein 2

GAPDH - Glyceraldehyde-3-phosphate dehydrogenase

GEx - Genome expression analysis

GRB7 - Growth factor receptor-bound protein 7

H3K4 - Lysine 4 on histone 3

Ha-RAS - Harvey rat sarcoma viral oncogene homolog

HCC - Hepatocellular carcinoma

HOXC8 - Homeobox C8

HSRs - Homogenously staining chromosome regions

HUVEC - Human umbilical vein endothelial cells

IP_3R - Inositol 1,4,5-triphosphate receptors

kDa - kilo Dalton

LEP - Luminal epithelial cells

LETM1 - Leucine zipper-EF-hand containing transmembrane protein 1

LETM2 - Leucine zipper-EF-hand containing transmembrane protein 2

LPA - Lysophosphatidate

LSM1 - LSM1 homolog U6 small nuclear RNA associated (S. cerevisiae)

MDa - Mega Dalton

MEP - Myo epithelial cells

miRNA - micro RNA

MLL1 - Mixed Lineage Leukemia 1

MYC - V-myc myelocytomatosis viral oncogene homolog (avian)

ncRNA - non coding RNA

NEM - *n*-ethyl maleimide

NHEJ - Non-homologous end joining

PA - Phosphatidate

PAP1 - phosphatidic acid phosphatase type 1

PAP2 - phosphatidic acid phosphatase type 2

PAT1B - Protein associated with topoisomerase II homolog 1 (S. cerevisiae)

P-bodies - Processing bodies

PgR - Progesterone receptor

PLK1 - Polo like kinase 1

PPAPDC1B - Phosphatidic acid phosphatase type 2 domain containing 1B

PPM1D - Protein phosphatase Mg^{2+}/Mn^{2+} dependent, 1D

PROSC - Proline synthetase co-transcribed homolog (bacteria)

RAB11FIP1 - RAB11 family interacting protein 1

RbBP5 - Retinoblastoma binding protein 5

REF - Rat embryonic fibroblast

RPS6KB2 - Ribosomal protein S6 kinase, 70kDa, polypeptide 2

SCRU - The Stem Cell Research Unit

shRNA - short hairpin RNA

siRNA - small interfering RNA

snRNA - small nuclear RNA

SNP - Single nucleotide polymorphism

SPFH domain - Stomatin/flotillin/HflK/C domain also known as Prohibitin (PHB) domain

sRNA - small RNA

STxB - Shiga toxin B

TfR - Transferrin receptor

TGOLN2 - Trans Golgi network protein 2

VAMP4 - Vesicle associated membrane protein 4

WDR5 - WD repeat domain 5

WHSC1L1 - Wolf- Hirschhorn syndrome candidate 1 like 1

WRAD - Wdr5 - RbBP5 – Ash2l complex

XNR1 - 5'→3' exoribonuclease

ZNF217 - Zinc finger protein 217

ZNF703 - Zinc finger protein 703

1. Introduction

1.1. Breast cancer in general

Breast cancer is the most common cancer among women in modern days and is additionally the leading cause of cancer deaths in women in the world (1). Both breast cancer incidence and mortality rates are higher in developed countries than in developing countries, probably due to different lifestyles and environment. Many factors are known to have an impact on cancer development, both genetic and environmental factors.

The largest risk factor for getting breast cancer is being a woman. There after comes family history. Individuals that have a first degree relative diagnosed with breast cancer have a two-fold increased risk of getting the disease as well. The risk is also significantly increased if a second or third degree relative is diagnosed (2). These factors are strongly related to another known risk factor, which is ethnicity, but breast cancer incidences are not only unevenly distributed globally but also ethnically, within countries (1, 2). Other factors that increase the risk of getting breast cancer or any cancer for that matter are environmental factors. What we eat and drink can have large impacts on our risks of developing cancer. Obesity, hormonal use, old age at first pregnancy and short duration of breast feeding can all contribute to the risk of developing breast cancer (2).

Even though all breast cancer incidents occur in the same organ the concept describes a wide range of tumors. Cancers are believed to start with a genomic alteration that gives the cell advantage that leads to more proliferation or avoidance of apoptosis (3). These alterations can rise from different start points. In families where breast cancer incidences are high the start point is likely a germ-line mutation or variation. Whereas in other cases the start point can be a somatic mutation that occurs in one cell at some point and increases the odds of other genomic mutations. The start points in somatic tumors are rarely known, for when cancer is diagnosed the tumors have often grown to be a mass of millions of cells. However for the familial breast cancer few mutations are known to increase the risk of developing tumors.

1.2. Familial breast cancer

History of breast cancer in the family is a well known risk factor. 15-20% of all breast cancer incidences are due to the genetic background of the patient (4). There of about 16-20% can be explained with mutations in the BRCA1 or BRCA2 genes (4, 5). The BRCA1 and BRCA2 genes are genes that show high penetrance when mutated (2, 4, 6). This means that an individual that carries a deleterious mutation within these genes increases his/her risk of developing cancer up to 50-80% life time risk (2). Other known predisposing breast cancer genes are e.g. CHEK2 (found in ~1% of the population in the United Kingdom) and BRIP1 (found in less than 0.1% of the population) (2, 4). There are also known a number of common low penetrant variants that affect breast cancer risk. Examples are variants within CASP8, CASP10 and FGFR2 (2, 4). Most of these low risk variants are not within a gene and the function is unknown.

Most breast cancer incidences are triggered by an external factor, even though some individuals are more prone to those external factors than others, such as carriers of germ-line mutations. Many genomic variants can increase the sensitivity but ultimately the environment or lifestyle push carcinogenesis over the start point. If external factors did not play this large role in tumor formation somatic alterations of the genome would occur to much less extent or not occur at all.

1.3. Breast cancer and genomic alterations

Majority of breast cancer incidences occur in persons that do not show family history of breast cancer, these cancers most likely arise from somatic genomic alterations. Alterations in genes that can be found in tumors are of various origins. Point mutations can occur within genes or their regulation sequences and disrupt the genes' function, leading to tumor formation. The larger alterations such as translocations and inversions are also disruptive. Inversions can destroy genes' function leading to functional loss (5, 7). Translocations can happen within a gene and therefore lead to loss of function but it can also occur between two genes leading to formation of fusion genes. An example of a fusion gene is when ABL and BCR translocate together to form the ABL-BCR fusion gene. This event is highly common in leukemia and is found in all cases of chronic myelogeneous leukemias (5). After the first step more genomic changes accumulate in the cancer, giving the cancer cells more and more advantage in each step, which eventually allows them to take over the tissue environment or even move to different tissues (3, 5, 8, 9).

Genomic alterations that seem to be fairly frequent in breast cancer are deletions and amplifications (7). When a genomic region is deleted the genes located on that particular region are lost and only one copy left for expression. Some chromosomal regions are more prone for deletions and these regions often carry tumor suppressor genes (5). On the other hand when a chromosomal region is amplified possible oncogenes are located within the region. An oncogene regulates cell proliferation and differentiation under normal circumstances, but can lead to tumor formation if the genes' function is disrupted. These genes can gain function when they are amplified (7, 10). The mechanism causing amplifications will be discussed here below.

1.3.1 Amplifications in breast cancer

Amplifications can be subdivided into two categories, a) double minutes (DMs), which are extra chromosomal copies or b) homogenously staining chromosome regions (HSRs) (11, 12). HSRs can be found as repeats within the chromosome it is derived from, or distributed in the genome (11). The initiation of amplification is the formation of breakage-fusion-bridges (BFB). These molecular bridges form as a result of double strand breaks that occur when telomeres are dysfunctional or when a break occurs at fragile sites in the DNA. The BFBs link together chromosomal regions by non-homologous end joining (NHEJ). Repairing chromosomes by NHEJ can then lead to dicentric chromosomes containing DNA repeats where their fusion occurred. As a result of anaphase separation of chromosomes the cell again recognizes the broken chromosome and BFB formation is often repeated cell cycle after cell cycle (12). DNA replication dysfunction has also been implicated as an initial

mechanism for amplification formation. When replication forks collapse the replication is initiated within the chromosomal region resulting in repeats of the region (11, 12).

When chromosomal regions are amplified the genes located within that region are frequently overexpressed as a result (10, 13). Certain chromosome regions are found more frequently amplified in cancer than others. Therefore these regions are thought to contain genes that in one way or another give the tumor cells an advantage to prosper or survive or both. Such genes are defined as oncogenes. An oncogene gives the cancer cells ability to grow faster, avoid apoptosis, induce angiogenesis or somehow gain advantage over other normal cells in that microenvironment (3, 10). The cells overexpressing an oncogene have gained advantage over the cells surrounding them, leading to evolutionary selection for cells carrying amplification (10-12).

Through the decades, since DNA amplification was described, evolution of methods has processed fast. There are several ways to estimate DNA levels in tumor samples. Fluorescent in situ hybridization (FISH) has been used to compare counts of centromere signals versus the signals given by amplified genes (11). This method has proven good to estimate translocations simultaneously to amplification, when the target gene is known (14). Methods like comparative genomic hybridization (CGH) have been used on microarrays (aCGH) (11). This method gives the possibility to compare the whole genome of tumor sample to normal DNA abundance (15, 16). By using aCGH unknown amplified regions have been identified and the size and amplitude of amplicons has been estimated (10). Further developments of methods for this purpose has generated analysis of single nucleotide polymorphism (SNPs) frequency as well as high throughput sequencing (11, 12, 17).

1.3.2. Identification of amplicons' target genes

As mentioned before genes that are amplified often show increased expression in tumors (10, 13). The expression of the amplified region can be measured to identify the target gene of each amplicon (10). By doing hybridization of isolated mRNA (converted into cDNA) on microarrays containing probes for each gene of the genome one can estimate the mRNA expression in tumors relative to expression in a normal tissue. The measurements of gene expression in breast tumors carrying amplification are important in order to find what genes pull the tumor cells through selection (8, 12). These methods, microarray based genome expression analysis (aGEx) are also used when identifying breast cancer subtypes, which will be discussed later in this thesis (18-20).

Increased expression of a gene simultaneous to its amplification is critical if the gene is to be considered a candidate driver gene of a specific amplified region (10). Genes that are amplified without increased expression are not likely to be a target of the specific amplification, but are dismissed as passenger genes (12). Proteins that are found in elevated levels when their transcribing genes are amplified can be classified as possible oncoproteins and their corresponding genes are possible oncogenes (10).

Previous studies done to identify candidate oncogenes within various regions in the genome included testing the genes' ability to transform normal cells (21). Transformation of cells means that a

cell line that cannot grow without the support of basal membrane gains the ability to form colonies in soft agar when an oncogene is overexpressed in the cell line (21). Soft agar which is used for these tests contains no imitating supporting features. The soft agar test is also called anchorage independent test. These methods have been used to study various genes that show oncogenic potential, such as amplification and overexpression, in tumors. If a gene plays a role in transforming cells, it is likely to contribute to factors such as cell proliferation and differentiation. In addition a gene that gives a normal cell line transforming abilities when overexpressed is likely to play similar roles in cancer cells that already overexpress the gene. Therefore tests for potential oncogenes also aim to study effects of knocking down potential driver genes in cell lines carrying amplification (10). In this research project we used the latter method for identification of potential oncogenes amplified and overexpressed in breast tumors.

1.3.3. Commonly amplified regions in breast cancer

Many chromosomal regions are known to be found amplified in breast cancer, the most commonly amplified regions are 1q, 8p, 8q, 11q, 12q, 17q, 20q (7, 8, 13, 15, 22, 23). For many of these regions, the target genes have been identified. Probably the most studied region is 17q, where the ERBB2 gene is localized. ERBB2 is overexpressed when amplified in breast cancer (24, 25). On chromosome 20q13 the target gene has been proposed to be ZNF217. This gene codes for Zinc finger protein 217, a transcription factor that regulates ERBB3 expression (26). Chromosomal region 11q13 is also frequently found amplified in breast cancer. Within the amplified region lies the gene that codes for cyclin D1 (CCND1) which is believed to be the target gene (27-29). Even though all these regions have been well characterized and their candidate genes have been identified research focusing on these regions is still identifying new putative oncogenes. For example the GRB7 gene and miRNA gene localized within the 17q amplicon have been shown to play a role in tumor formation along with ERBB2 (25, 30). On chromosomal region 11q the GAB2 has also been identified as a candidate driver gene of the amplicon (29). These results indicate that within each amplicon there might be more than one oncogene, and that each gene probably plays a distinctive role in the environment it exists. In addition different amplicons are often amplified simultaneously. For example, chromosome 17q is often found amplified together with 11q13 and chromosome 20q13, and the amplicon on 8q24 is frequently found amplified together with 11q13 and 20q13 as well (22, 24). This implicates further that within these amplicons lie genes that play different roles depending on the co-players (that is to say co-amplification).

1.4. Breast cancer subtypes

Breast cancer is a common name for a disease of various origins. The tumors can be classified in numerous ways, for example pathological classifications distinguish between tumors by estimating the origin of the tumor within the organ, lobular or ductal origin. Another way to classify tumors is by tumor size, histological grade and lymph node invasion. In addition tumor samples often undergo staining for various membrane bound markers, like hormone receptors (31). The past decade further classifications have been emerging. These classifications are based on the molecular type of the

tumors. The tumors are divided into subgroups based on their mRNA or protein expression of certain molecules tested (18, 19, 32, 33)

Sørlie et al. classified breast cancer into five molecular based subtypes: ERBB2 overexpressing tumors, luminal A, luminal B, basal and normal like breast tumors. The subtypes were classified based on the mRNA expression of the whole genome (18, 19). Each subtype displays different expression of groups of genes that can distinguish the subtypes (18-20). These observations have also been verified with immunohistochemistry in both breast tumors and breast cancer cell lines (32, 34). The subtypes differ in various ways but they carry a few trademarks that will be listed here. The ERBB2 subtype counts for tumors overexpressing the ERBB2 gene, luminal A subtypes is estrogen receptor (ER) and progesterone receptor (PgR) positive, luminal B is also ER and PgR positive but in addition expresses the ERBB2, and is also a more aggressive tumor type, compared to the luminal A group. The basal subtype is sometimes called triple negative because these tumors stain negative for ER, PgR and ErbB2. And at last there is the normal like subtype that resembles normal tissue (18-20, 35).

Recent studies have changed our knowledge of breast cancer subtypes adding to the subgroups as well as questioned the existence of defined subgroups (36). Now the subtypes are considered to be: the ERBB2 overexpressing subtype still counts for tumors that carry amplification of the chromosomal region carrying ERBB2 gene (36). Luminal A: a subgroup that is among other ER and PgR positive, Luminal B: a subgroup that is also ER and PgR positive but shows higher Ki-67 expression and proliferation. Some define the third Luminal group, which is ER, PgR and ErbB2 positive, sometimes called a Luminal/Her2 subtype (36, 37). In fact not everyone agrees that the luminal subgroup can be divided into any subcategories, because they are all rather similar (36). The remaining tumors are triple negative that means they do not stain positive for ER, PgR or ErbB2. Until recently this group was divided into basal and normal-like tumors. A study done by Prat et al. identified a new subtype, namely the claudin low subtype (38). These tumors, in addition to being triple negative, express immune response genes, carry stem cell like features and are prone to undergo epithelial-to-mesenchymal transition (EMT). It is now widely accepted that the normal-like tumors are an artifact but not an actual subtype of tumors. The triple negative tumors are therefore divided into the claudin-low subtype and the basal subtype, that displays expression of cytokeratins 5 and 17 as well as CD44 and EGFR, for example (31, 36). Further studies will be needed to declare if the triple negative subgroup consists of more subtypes.

This molecular classification is important for determination of treatment. The overall survival differs between subtypes, in ER positive tumors, that are luminal A and luminal B the mortality rate is higher for the luminal B subtype than the luminal A. The worst overall survival is found within the basal subtype and the ERBB2 overexpressing tumors (15, 18-20, 32, 35). In addition the metastatic behavior of tumors differs between subtypes. Basal tumors tend to metastasize to brain and lung while luminal subtypes and ErbB2 overexpressing tumors tend to invade bone tissue. These different metastasis patterns are reflected in relapse and overall survival (37).

These different subtypes all harbor different genomic alterations, ERBB2 overexpressing subtype, like the name indicates carries amplification on the 17q region, resulting in overexpression of the ERBB2 gene as well as other genes located within that amplicon (7, 15). Luminal subtypes carry high amplifications, most often found on chromosomal regions 8p11-p12 and 11q12. Luminal A tumors in addition carry few copy number change, often they carry 1q and 16p gain as well as 16q loss. The luminal B subtype on the other hand is more complicated with more high level amplifications on a variety of chromosomal regions, like gains on 17q and 20q, along with loss on 17p and 22q (7). The basal like subgroup carries multiple low level alterations and only one or two high level amplification, sometimes found on chromosome 8q24 where MYC is localized and on 17q23 where the PPM1D gene is found (7). As reviewed above tumors rise from different mechanisms: point mutations, deletions or amplifications for example. The cause and result of these mechanisms and their link to different subtypes has not been determined even though studies have shown that some alterations are more common in some subtypes than others (15). Amplifications, which are quite common in breast cancer, have been linked to few different subtypes, like the luminal subtypes as well as ErbB2 overexpressing subtype (7).

An aCGH performed on Icelandic breast tumors revealed, among other things, amplification on chromosomal regions 17q12, 11q13 and 8p12 in the frequency of 15%, 14% and 14%, respectively. The chromosomal region 8p12 was chosen for further studies (16). The project represented in this thesis is a continuation of the project described in (16) and based on its results.

1.5. Chromosomal region 8p12-p11

The chromosomal region 8p12-p11 has been reported to be amplified in 10-23% breast cancer cases (16, 22, 39-41). No international standard has been created to distinguish between gain of chromosomal regions and amplification, and that might at least in part explain why such variety is found in the frequency of amplification on the 8p12-p11 region. In addition studies are performed in combination of differently originated groups of tumors. Some studies have shown that amplification on 8p12-p11 is associated with poor clinical outcome (22, 23, 42). These observations have also been associated with a particular gene on the 8p12-p11 region, such as ZNF703 (40, 43). In addition other studies have shown that amplification on chromosomal region 8p12-p11 in combination with amplification on 11q13 have more impact on patients outcome than amplification on only one of the two loci (28, 44). In fact 8p12-p11 is often co-amplified with the amplicon on 11q13, which suggests some kind of cooperation between genes located on these regions (7, 8, 15, 22, 27, 28). In addition the amplicon on chromosome 8q24, 12p13-15 and 20q13 are also commonly co-amplified with 8p12-p11 (15, 22).

Several groups have made an effort to identify the target gene within the amplicon 8p12-p11 (27, 39-41, 43, 45-47). Previous studies in our lab have shown that the minimal region of amplification counts 1.6 Mb region spanning from 37.3 - 38.9 Mb and within this region 20 genes are localized (16). Another study, done by Gelsi-Boyer et al., 2005, identified the chromosomal region from 35.97 Mb to 42.33 Mb as the 8p12-p11 amplicon (42). Within this region amplicon cores have been found, which

means that some genes are probably more important for the cancer progress than others (10). It is highly likely that some genes are only amplified because they are located close to the gene that is the main target of the amplification (11). Since at least four different amplicon cores have been identified within the 8p12-p11 amplicon, it could be postulated that each core harbors at least one gene that stimulates in some way carcinogenesis (42).

Correlation between DNA amplification and mRNA as well as protein expression was tested, in order to identify candidate target genes of the amplification. For these studies aCGH, aGEx and western blot were used, for DNA, mRNA and protein levels measurements, respectively (16). For genes to be listed as a candidate target gene they had to fulfill four criteria. In breast tumors 1) the gene had to be overexpressed on mRNA level when amplified, 2) positive correlation had to be observed between mRNA and protein levels 3) as well as DNA copy number and 4) the genes protein product had to be more abundant in three out of four breast cancer cell lines carrying amplification on this regions than breast cancer cell lines without amplification (16). Seven genes met these criteria and therefore were identified as candidate target genes of the amplicon. Eleven genes were eliminated based on lack of correlation with mRNA or protein expression in breast tumors or breast cancer cell lines, and two genes were not tested because of lack of experimental tools (16). Those two genes, LETM2 and PPAPDC1B will be described below.

1.5.1. LETM2

Leucine zipper-EF-hand containing transmembrane protein 2 (LETM2) codes for a protein that has not been studied in detail, however it resembles another transmembrane protein called LETM1 (48). Both proteins are localized in the mitochondria membrane, but Letm2 is thought to have one hydrophobic region that sits in the membrane (48). Letm1 helps maintain mitochondria structure and ion concentrations in the mitochondria (48, 49). When Letm1 was downregulated in cell lines mitochondria swelling was observed, as well as fewer cristae appeared to form in cells lacking Letm1, and the cristae that did form seemed to be disorganized, compared to cells expressing Letm1. In addition to this, cells lacking Letm1 could not assemble the respiratory chain properly; the protein complexes that form proton pumps in the mitochondria membrane were not able to be formed, even though the protein's abundance did not differ from normal conditions (48). The Letm1 protein has also been implied in Ca^{2+} uptake, in mitochondria, but in cells where Letm1 was downregulated cytosolic Ca^{2+} uptake was diminished (49). Based on the similarities between these proteins, sequence and structure, and their localization in the cell, Letm2 could possibly play a role similar to Letm1. In contrast overexpression of Letm2 in cells where Letm1 has been downregulated did not result in normal mitochondria structure (48).

Even though Letm2 is localized in mitochondria like Letm1 the expression patterns in various tissues differ between the proteins. While Letm1 showed protein expression in all rat tissue tested the Letm2 protein only gave a strong signal in rat testis (48). When tissue sections of rat testis were stained with Letm2 antibody the protein stained positive in spermatogenic cells and spermatozoa. Therefore the protein probably plays a role in spermatozoa development (48).

LETM2 has been found amplified and sometimes the mRNA is found overexpressed as well in breast cancer (16, 41, 43). Probably due to lack of information about the protein little work has been done in order to measure the protein expression in cancer. One study has been published where the protein was overexpressed in MCF10A cell line. MCF10A is a cell line that has been immortalized but not transformed, so the cell line is able to grow infinitely but does not form colonies in soft agar (50). When LETM2 was overexpressed in MCF10A no effects of the cells ability to form colonies was observed (45). Without further knowledge of Letm2 function, speculations about the proteins role in carcinogenesis might be difficult.

1.5.2. PPAPDC1B

Phosphatidic acid phosphatase type 2 domain containing 1B (PPAPDC1B) is a gene that codes for a protein that plays a role in lipid metabolism (51-53). The protein is a Mg^{2+} independent enzyme that resembles yeast diacylglycerol pyrophosphate phosphatase (DPP1), therefore the protein has also been called diacylglycerol pyrophosphate phosphatase like 1 (DPPL1) (53). Phosphatases have been classified into two distinct groups, phosphatidic acid phosphatase type 1 and type 2 (PAP1 or PAP2). They distinguish in their dependency on Mg^{2+} and their sensitivity to *n*-ethyl maleimide (NEM), where PAP1 is Mg^{2+}-dependant and NEM-sensitive but PAP2 Mg^{2+}-independent and NEM-insensitive (54, 55). The Ppapdc1B protein seems to be some kind of intermediate because the protein is Mg^{2+}-independent and NEM-sensitive (53). The phosphatases role is to dephosphorylate different substrates to yield phospholipids for both membrane synthesis and signaling molecules. The proteins are membrane bound; they contain transmembrane domains and are found in various membranes, including the Golgi system and in some cases the plasma membrane (51, 52, 54). Ppapdc1B dephosphorylates substrates like diacylglycerol pyrophosphate (DGPP), phosphatidate (PA) and lysophosphatidate (LPA), mainly PA and LPA since DGPP is not found in mammals (53).

The gene has been implicated as an oncogene in breast cancer but as a metastatic suppressor gene in hepatocellular carcinoma (HCC) (39, 56, 57). In breast cancer the chromosomal region harboring the PPAPDC1B gene is often amplified, and as a result the mRNA is overexpressed (39, 41, 43). Bernard-Pierrot and her colleagues used siRNA to downregulate PPAPDC1B mRNA expression in breast cancer cell lines, harboring amplification of the gene. The downregulation led to less cell proliferation, induced apoptosis, decreased focus formation and decreased growth in soft agar. When the cells were transplanted into mice, the downregulation resulted in decreased tumor growth (39). These results indicate that Ppapdc1B might be important for tumor growth both in vitro and in vivo.

The Ppapdc1B protein has also been studied in HCC, where the protein seems to function as metastatic suppressor protein (56, 57). In HCC metastasis the chromosomal region carrying PPAPDC1B is often lost. Downregulation of the gene in HCC cell lines that had high metastatic ability inhibited the cells invasiveness, even though the downregulation did not affect the cells proliferation (57). The Ppapdc1B protein has been found in gradient expression from metastasis to tumor to normal tissue, where the expression is highest in normal tissue (56). In addition the protein was found to

associate with increased overall survival as well as decrease in lung metastasis in mice (56, 57). Taken together the results of these studies in both breast cancer and HCC indicate that the Ppapdc1b protein might function differently in these two tissues and that within these cancer types there are different mechanisms driving the cancer growth.

1.5.3. Candidate genes within the 8p12-p11 amplicon

As described above the studies of Einarsdóttir et al. identified seven genes located within the 8p12-p11 amplicon as putative target genes of the amplification and as putative oncogenes (16). These seven genes code for proteins playing different roles in the cell, some regulate gene expression, others cellular trafficking or protein degradation and the function of some of the genes function is still relatively unknown, as described below.

1.5.3.1. ERLIN2

ERLIN2 is one of the candidate genes localized within the 8p12-p11 amplicon. This gene showed good correlation between DNA amplification, mRNA expression as well as protein expression (16). It codes for a protein called SPFH domain containing protein 2 (SPFH2) or Endoplasmic reticulum lipid raft protein 2/ER lipid raft associated 2 (Erlin2) (58-60). In this thesis the protein will be named Erlin2. The protein was first identified as a marker for lipid rafts in inner cell membranes that were not considered to contain any lipid rafts due to their low cholesterol and sphingolipids concentrations (58, 60, 61). The protein resides in the endoplasmic reticulum (ER), including the outer nuclear membrane. A large part of the protein is located in the ER-lumen (endoplasmic reticulum-lumen), while the protein is anchored to the ER membrane on the N-terminus (58-60, 62). One study has also identified the protein in low concentrations in the cells plasma membrane in human umbilical vein endothelial cells (HUVEC) (60).

Erlin2 forms homo- and hetero- oligomers with a similar protein called Erlin1 and together the proteins also form high molecular weight complexes that are thought to be somewhere between one and two MDa in size (62-64). The Erlin2 and Erlin1 complexes have been shown to form in 2:1 ratio, respectively. Erlin1 and Erlin2 are highly similar proteins, with 89% amino acid similarity, but they mostly distinguish from each other on the C-terminal region, which seems to be the region necessary for these protein-protein interactions (58, 63). Despite these similarities Erlin2 deficiency cannot be rescued by increased Erlin1 abundance but Erlin2 can rescue Erlin1 deficiency in rat-cells (62).

The Erlin2 protein has recently been shown to play a role in the ER-associated degradation (ERAD) pathway (59, 62, 64). The Erlin2 and Erlin1 complex mark Inositol 1,4,5-triphosphate receptors (IP_3R) for ubiquitin mediated degradation, after the receptors' activation. This has been shown in both rat and human cell lines (59, 62, 64). IP_3R play an important role in mediating apoptosis (65-67). In mammals three different types of IP_3R exist, IP_3 receptor 1 (IP_3R1), IP_3 receptor 2 (IP_3R2) and IP_3 receptor 3 (IP_3R3) (65, 68). All three receptors are likely target proteins of ubiquitination mediated by the Erlin2/Erlin1 complex, and studies have shown that this is at least the case for IP_3 R1 in human cell line HeLa (64). The formation of the Erlin2/Erlin1 complex is necessary for their function.

When residues responsible for the complex formation are truncated, degradation of IP_3R does not occur (62, 64).

Studies have shown that if Erlin2 is downregulated the ubiquitination of IP_3R decreases however when Erlin1 is downregulated Erlin2 seems to rescue the deficiency and the complex' function increases, resulting in increased ubiquitination of IP_3R (62). Studies on T-cells have shown that T-cells lacking IP_3R, (IP_3R1 completely and IP_3R2 and IP_3R3 at least partly) can avoid undergoing apoptosis and therefore survive chromosomal aberrations that internal cell systems should identify as death signal (68). Overexpression of Erlin2 as a result of DNA amplification might be able to deprive the cells of its IP_3R, giving the cells the ability to avoid apotosis.

This protein has not been studied much in regards of its role in breast cancer, but few groups have described the gene's amplification and overexpression in breast cancer (27, 41, 43). One article has been published where the protein was overexpressed in MCF10A cell line (45). There the overexpression resulted in colony formation when cells were cultured in insulin free media. The cell proliferation also increased modestly (45).

1.5.3.2. LSM1

LSM1 codes for a protein called LSM1 homolog U6 small nuclear RNA associated (S. cerevisiae) (LSm1). The protein is well conserved through evolution and has been identified in both eukaryotes and Archaea (69, 70). The protein contains a Sm domain that plays a role in mRNA processing, proteins containing this particular domain are found both in the nucleus and cytoplasm, and in both cases bound to RNA molecules (70-72). The LSm proteins have mostly been studied in yeast but recent studies show that yeast and human LSm proteins interact with similar proteins where they play the same role (73-76). The LSm1 protein interacts with six other LSm proteins, from 2 to 7, to form a protein complex that localizes in cytoplasmic foci or so called Processing bodies (P-bodies) (73, 76). In contrast of the other six LSm proteins in the complex, LSm1 is the only one that is only found in the cytoplasm, not in the nucleus as well. In addition the LSm1 protein is found bound to RNA molecules such as U6 small nuclear RNA, this indicates that LSm1 does not play a role in pre-mRNA splicing (71, 77).

LSm1-7 forms a ring shaped complex which can both interact with other proteins as well as binding to RNA molecules (71, 72, 75). In human this protein complex interacts with a number of proteins that participate in mRNA degradation, like Protein associated with topoisomerase II homolog 1 (S. cerevisiae) (Pat1b), DCP1 decapping enzyme homolog A (S. cerevisiae) (Dcp1a), DCP2 decapping enzyme homolog 2 (S. cerevisiae) (Dcp2) and 5'→3' exoribonuclease (Xnr1), and more interacting proteins are probably yet to be identified (73, 75, 76). Together these proteins regulate mRNA abundance in the cytoplasm, by decapping and deadenylating the mRNA and as a result the molecule is degraded in the P-bodies (73-76). LSm1 therefore is crucial for the cell to regulate gene expression, through mRNA decay. Studies done in yeast indicate that all mRNA degradation is regulated via LSm1-7/Pat1 pathway, but there is evidence to support that this system is more complicated in human cells, whereas at least Pat1b seems to be gene specific (74).

LSm1 overexpression has been implicated with breast cancer and lung cancer (39, 41, 78). In all cases the gene is overexpressed, when mRNA levels are estimated and protein levels seem to correlate with mRNA overexpression in those cases where protein expression was measured (16, 78). Studies where LSM1 has been downregulated, in cells overexpressing the gene, or LSM1 overexpression is mediated in cells carrying normal DNA levels, indicate that the gene could have effects on carcinogenesis. After downregulation of LSM1 the cells proliferation decreased and apoptosis seemed to increase or cells lost the ability to grow in soft agar (39, 46, 78). In agreement with these results cell proliferation in MCF10A cells where LSM1 overexpression was mediated seemed to increase and the cells gained the ability to grow in soft agar, either in cooperation with other genes from the 8p12-p11 amplicon or when the cells were deprived of insulin from the medium (41, 46). Since LSm1 plays an important role in mRNA processing and degradation one could speculate that altering LSm1 gene expression might affect various pathways that involve proteins coded for by LSm1s target genes.

1.5.3.3. ASH2L

ASH2L codes for a trithorax protein called Ash2 (absent, small or homeotic) - like (Drosophila) (Ash2l). The gene was first discovered in Drosophila as a gene that was expressed during development of the larvae and caused death when mutated, in a large lethal mutations screening (79). Later the gene was identified and cloned in human, where it is expressed in various tissues primarily in heart and testis of an adult or lung and liver in a fetus (80). The gene is highly conserved through evolution and is for example found in the fruit fly, mouse and yeast (80).

The protein, Ash2l, binds to WD repeat domain 5 (Wdr5), Retinoblastoma binding protein 5 (RbBP5) and Dpy-30 homolog (C. elegans) (Dpy-30) (81-84). This protein complex has been called WRAD, a complex that often interacts with Mixed Lineage Leukemia 1 (MLL1) complex (81, 84). These protein complexes are important for methylation of lysine 4 on histone 3 (H3K4) (82, 83). In mammals approximately ten H3K4 methyltransferases are predicted (85). At least four methyltransferases are of the kinds that contain Ash2l. These complexes are called SET1-like complexes for they resemble the Set1 complex in yeast (82). Ash2l has been linked to trimethylation of H3K4, but the chromatin is thought to have undergone mono- and di- methylation, which is mediated via MLL complexes harboring at least RbBP5 and Wdr5 (82). Trimethylation on H3K4 is a marker of active chromatin (85). The complexes that Ash2l forms have so far only been implicated to transfer methyl to H3K4 (81-84). Defining the roles of each protein during methylation has not been completed, but studies show that Ash2l binds to Wdr5 and RbBP5 for methylation (81-84). Dpy-30 has newly been identified as a unit for stabilization and specification in the WRAD complex (81).

Ash2l has been found to bind to activating protein 2 δ (Ap2δ) in mammalian cells. Ap2δ is a transcription factor and when Ash2l interacts with this protein transcription activity of Ap2δ is increased (86). The binding of Ash2l and Ap2δ is associated with MLL2, so the complex probably plays the same role as the WRAD complex, in methylation (86). Another role proposed for Ash2l is in endosomal trafficking where components of the WRAD complex, such as Dpy-30 shows a signal in the Golgi apparatus when cells are stained for the protein. Neither Ash2l or RbBP5 show localization elsewhere

but in nucleus, however knock down of one of these three proteins results in a dysfunction of the endosomal trafficking in HeLa cells (87).

Regarding the proteins role in carcinogenesis few studies have been done where this protein plays a lead-role (88). The protein has been found to be increased in various cancer types, even if mRNA levels do not seem to increase (88). In breast cancer the gene is often amplified and as a result overexpressed both at mRNA and protein levels (16, 27, 39, 41, 43). In two studies where the protein is overexpressed in cell lines the protein affects cells carcinogenesis (45, 88). When the gene is transfected into rat embryo fibroblasts (REF) along with Ha-RAS the cells formed foci that REF cells do not normally do, and these same cells could form tumors in rats when injected (88). In agreement with these findings cell proliferation slightly increased when the gene was overexpressed in MCF10A as well as the cells formed expanded colonies when cultured in insulin free media (45). Ash2l regulates the gene expression of a number of genes. One of those genes is Homeobox C8 (HOXC8) which is an important gene in development (82, 86). Genes that are involved in development also often contribute to cells growth and differentiation. These features are important for oncogenes as well (3) So Ash2l might play a role in carcinogenesis by increasing expression of genes that could influence developmental pathways.

1.5.3.4. PROSC

PROSC was cloned in 1999 and codes for a poorly understood protein called proline synthetase co-transcribed homolog (bacteria) (89). Since the gene was cloned no studies have been done in order to understand and characterize the gene or its protein product (89). This gene has however often been found amplified in breast cancer, where mRNA levels are increased as well (16, 27, 39, 41, 43). In addition we observed increased protein expression in breast tumors harboring DNA amplification on the 8p12-p11 region (16). This gene´s oncogenic ability has been tested by overexpressing the protein in MCF10A breast cells. When Yang and his colleagues overexpressed PROSC in this cell line, it resulted in the formation of expanded colonies and slightly increased cell proliferation when cultured in insulin free medium. However they did not observe any growth in soft agar (45).

The lack of knowledge about this protein makes it both interesting and challenging for studies in breast cancer cell lines. Any effects that might be observed could hopefully clarify the protein's role in human cells.

1.5.4. Additional candidate genes within the 8p12-p11 amplicon

The four genes listed above were studied in various breast cancer cell lines and the results from those experiments are described in this thesis. Three candidate genes within the 8p12-p11 were not included in this M.Sc. project. They are BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like (BRF2), RAB11 family interacting protein 1 (class I) (RAB11FIP1) and DDHD domain containing 2 (DDHD2) (16). Brf2 is a subunit of a transcription complex in the cell, regulating expression of specific genes (90). Rab11fip1 regulates endocytosis and plasma membrane recycling (91, 92). Ddhd2 is a phospholipase that regulates membrane trafficking from the Golgi apparatus to

the plasma membrane (93). All three proteins play important roles in the cell and disruption of their activity such as overexpression might result in carcinogenesis.

1.5.5. Identification of a putative oncogene within the 8p12-p11 amplicon

Previous studies in our laboratory have identified seven genes within the 8p12-p11 amplicon as potential targets of the amplification (16). The next step is to test these genes and estimate their ability to drive carcinogenesis or in other words their oncogenic properties. Many groups have been studying the genes located within the amplified region and the results from those studies show how complicated the 8p12-p11 amplicon story is (27, 39-41, 43, 45-47). Few studies have identified the same gene as the target gene and the results of some studies indicate that one overexpressed gene from the 8p12-p11 region is not sufficient to drive carcinogenesis (27, 41). In the M.Sc. project described in this thesis a number of breast cancer cell lines were used to study the effect on cell proliferation by siRNA mediated knock down of candidate target genes of the 8p12-p11 amplification.

2. Aims

The main aims of this study were two:

1. To identify candidate genes located on the amplified region of 8p12-p11 in breast cancer. Seven putative oncogenes had been identified in a previous MSc. project done in our lab (16). Two genes, LETM2 and PPAPDC1B, remained to be analyzed in regards of protein expression and its correlation to either gene copy number and/or mRNA expression. The results are presented in this thesis.
2. To test if selected candidate target genes of the amplification at 8p12-p11, would affect cell proliferation in breast cancer cell lines. Of the genes identified as putative oncogenes, four were selected for further studies. These genes were ERLIN2, LSM1, ASH2L and PROSC. Expression of the genes were downregulated in cell-lines carrying 8p12-p11 amplification and cell proliferation was estimated. In addition eventual cooperation of three of these genes (LSM1, ASH2L and PROSC) in cell proliferation was studied by simultaneous downregulation of their expression.

Since these methods had not been used in our lab previously another aim of the project was to optimize the setup of experimental conditions.

3. Materials and methods

3.1. Cell lines

For correlation studies of PPAPDC1B´s and LETM2´s gene copy number versus mRNA and protein expression the breast cancer cell lines were selected based on their DNA status alone. The DNA status is based on published data. Four carried amplification: SUM52, SUM44, CAMA-1 and ZR-75-1 (27, 41, 94-97). Two were defined carrying gain, SUM225 and HCC1500 (27, 41, 97). Four cell lines carry normal DNA copy number in regards of the two genes; SUM229 harbors a gain in chromosomal region 8p12-p11, but for the LETM2 and PPAPDC1B genes the DNA quantity is normal (41). MCF10A, originates from fibrocystic disease, established without viral or chemical interventions and is considered to be normal with respect to chromosomal region 8p12-p11 (50, 95, 98, 99). Myo-epithelial cells (MEP) and luminal epithelial cells (LEP) isolated from reduction mammoplasty as described by Gudjonsson et al. (100) are assumed normal. LEP was only used for analysis of LETM2. Protein extracts from both MEP and LEP were a kind gift from the lab of Þórarinn Guðjónsson and Magnús Karl Magnússon, the Stem Cell Research Unit (SCRU) at Læknagarður. We also used two cell lines that have lost a copy of the two genes, MCF7 and MDA-MB-231 (95, 98).

For estimation of protein expression in breast cancer cell lines, compared to normal expression, lysates from the D492 cell line were obtained as a kind gift from the laboratory of Þórarinn Guðjónsson and Magnús Karl Magnússon, SCRU at Læknagarður. D492 is a breast cell line isolated from reduction mammoplasty, purified and immortalized using E6/E7 genes from the human papilloma virus type 16 (100). The D492 cell line is considered to be a normal cell line, that is the cell line is immortalized but not transformed so it does not have the same carcinogenic abilities as a cancer cell line does (100). The cell line was cultured in a 3D culture where the cells formed branch like structures. These structures were lysed and protein extraction was performed. We refer to this lysate as D492-3D. Protein expression in the D492-3D cells was used as normal expression, based on that 3D matrigel culture mimics the human body conditions better than 2D cell culture.

Cell lines that were used for siRNA transfection were selected based on their DNA status and protein expression of the genes in question (Table 5).

3.1.1. Cell culture conditions

CAMA-1 was purchased from American Type Culture Collection (ATCC). The cell line was cultured in DMEM/F12 (Gibco-Invitrogen) which was supplemented with 10% Fetal bovine serum (FBS) (Gibco-Invitrogen) and 0.1 nM Non-Essential Amino Acids (NEAA) (Gibco-Invitrogen). MCF7 and T47D breast cancer cell lines were obtained from Dr. Jill Bargonetti-Chavarria at the Hunter College biology department at the City University of New York. The MCF7 cell line was cultured in MEM (Gibco-Invitrogen) supplemented with 10% FBS, 10 µg/ml insulin (Sigma-Aldrich) and 1 nM sodium pyruvate (Gibco-Invitrogen) and T47D was cultured in RPM1-1640 (Gibco-Invitrogen) where 10% FBS was added along with 9.1 µg/ml insulin. SUM52 was purchased from Asterand and cultured in DMEM/F12 cell culture medium supplemented with 5% FBS, 5 µg/ml insulin and 1 µg/ml hydrocortisone (HC)

(Sigma-Aldrich). ZR-75-1 cancer cell line was a gift from the laboratory of Þórarinn Guðjónsson and Magnús Karl Magnússon, SCRU at Læknagarður. The cell line was cultured in RPMI-1640 supplemented with 10% FBS. All cells were cultured at 37°C in a humidified incubator with 5% CO_2 as recommended. Cells medium was changed when needed. The cell culture medium contains a color, phenol-red that turns orange or yellow when pH changes occur in the medium; this was used to determine if medium changes were needed. Cells were split upon approximately 90% confluency.

3.1.2. Cells identification

To verify that our cells were identical to the distributors cell lines we analyzed genetic markers listed on the distributors' webpage www.atcc.org. The ATCC show DNA profiles for all cell lines available. This test was not done for SUM52, because the DNA profile for that cell line is not available and not for T47D, because this cell line was added late in the process of this thesis. DNA extraction from the cells was carried out using Wizard genomic DNA purification kit (Promega) according to manufactures protocol. The DNA profiling was done by Haukur Gunnarsson MSc. at our laboratory. For the profiling he used PowerPlex 16 HS system (Promega). This kit allows detection of 16 loci. For all cell lines tested the DNA profile matched the profile listed by the distributor, ATCC.

3.1.3. Contamination test

For all cell lines mycoplasma tests were carried out. DNA isolated from the cells, as described above, was used. In addition DNA was isolated from the cell culture medium by heating a portion of medium, collected from the cells, to 95°C for 5 minutes. The samples were then centrifuged and supernatant was collected for further tests. The samples were mixed for polymerase chain reaction (PCR): H_2O, 10x PCR Gold buffer (Roche), MgCl 2 mM (Roche), Nucleotide mixture 0.64 mM (Fermentas) and AmpliTaq Gold (Roche). The primers that were used for the reaction were specific for the 16s rDNA of mycoplasma purchased from Eurofins MWG Operon. (Forward primer: cgcctgagtagtacgtwcgctgcctgrgtagtacattcgccgcctgagtagtatgctcgccgcctgggtagtacattcgc, reverse primer: gcggtgtgtacaaracccgagcggtgtgtacaaaccccga, w = mixture of t and a, r = mixture of a and g). Both primers were added to the reaction in 0.2 mM concentrations. 50 ng of DNA were used for each PCR reaction. The samples then underwent the following PCR program:

95°C, 5 minutes

5 cycles (94°C, 30 sec, 50°C, 30 sec, 72°C, 1:10 min)

35 cycles (94°C, 15 sec, 56°C, 15 sec, 72°C, 55 sec)

72°C, 7 min

cooled down to 4°C.

DNA samples from all cell lines were used for the PCR tests, as well as negative control containing water and positive control containing mycoplasma DNA, which was a kind gift from Christian Praetorius at the Department of Biochemistry and Molecular Biology, faculty of Medicine, University of

Iceland. The PCR products were put on 1.5% agarose gel, containing ethidium bromide, and electrophoresis was run for approximately 40 minutes. The gel was then visualized on UV table, to examine if any PCR products were seen. Positive controls were visible as expected but all cell lines tested negative for mycoplasma.

3.2. siRNA transfection

siRNA transfections were carried out in 96- or 24-well plates for protein isolation and 96-well plates for cell proliferation tests. Cells were transfected when they were approximately 30-50% confluent, varying a little between cell lines. All transfection solutions were used at room temperature. For transfection Opti-MEM (Gibco-Invitrogen) cell culture medium was used for optimal knock down. The transfection reagent, oligofectamine (Sigma-Aldrich), was used for all cell lines to deliver the siRNA molecules into the cells. siRNA molecules were purchased and designed by Ambion – Applied Biosystem. All siRNA purchased were Silencer Select siRNA. The siRNA sequences and concentrations used are listed in Table 1.

Table 1. A list of siRNA used for the process of this thesis.

Target gene	**Exon**	**[siRNA] nM concentration used**					**Sequence 5'→3'**	
		CAMA-1	MCF7	SUM52	T47D	ZR-75-1	Sense	Antisense
ERLIN2	5	5 nM	5 nM	nd	nd	1 nM	GAAUUGAAGUGGUGAACUUtt	AAGUUCACCACUUCAAUUCtg
PROSC	5	2.5 nM	20 nM	nd	20 nM	nd	GGUUAAAGGUUAUGGUCCAtt	UGGACCAUAACCUUUAACCtt
	8	2.5 nM	20 nM	nd	20 nM	nd	GGAUCUACAAAUGUCCGAAtt	UUCGGACAUUUGUAGAUCCta
LSM1	2	5-10 nM	5 nM	nd	nd	5 nM	GCAUUGAUCAAUUUGVAAAtt	UUUGCAAAUUGAUCAAUGCtt
ASH2L	10	2.5-10 nM	2.5 nM	10 nM	nd	nd	GGAACACCCGUUUAACAAAtt	UUUGUUAAACGGGUGUUCCaa
	12	2.5-10 nM	2.5 nM	10 nM	nd	nd	GCUCCUUUAGGUUAUGAUAtt	UAUCAUAACCUAAAGGAGCtt
PPM1D	2	nd	nd	nd	nd	50 nM	GGAAAGAGAACGAAUCGAAtt	UUCGAUUCGUUCUCUUUCCtt
	6	nd	50 nM	nd	nd	50 nM	GGACAUUAGAAGAGUCCAAtt	UUGGACUCUUCUAAUGUCCtt
PLK1	5	10 nM	50 nM	10 nm	20 nM	50 nM	CCAUUAACGAGCUGCUUAAtt	UUAAGCAGCUCGUUAAUGGtt
No gene (siNEG)		2.5-10 nM	2.5-50 nM	10 nM	20 nM	1-50 nM	No sequence listed	No sequence listed

Nd = Not done.

The preparation of transfection solutions was as follows. The siRNA were suspended in RNase free water to make 1 µM siRNA solution. The siRNA solution was then suspended in Opti-MEM to yield desired concentrated solution for each gene and each cell (Table 1), this will be referred to as solution A. Solution B contained Opti-MEM and oligofectamine in the 4:1 ratio, respectively. After solution B had been mixed it was kept at room temperature for five minutes. Then solution A was added to solution B and the mixture was kept at room temperature for 20 minutes. During this time the oligofectamine and the siRNA formed molecules that can enter the cells without harming them. Afterwards the mixture was diluted into the wells. The cells were washed carefully with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2.0 mM KH_2PO_4) previous to adding the transfection mixture. 20-100 µl of the transfection mixture was added to each well, depending on the size of the wells. Four to six hours after transfection Opti-MEM was added to the cells, in the same amount as the transfection solutions. Cells were incubated overnight, in cell culture incubator and the next morning normal cell

culture medium replaced the transfection solutions. Cells were cultured for 72 – 120 hours and then cell counting and knock down estimations were performed.

All experiments were carried out in triplicate, except for the downregulation of PROSC, but that experiment was only successful one time, in each cell line. Various transfection times were chosen for each cell line and each gene for optimal transfection. In addition the cells had to undergo at least two cell divisions for effects on cell proliferation to be observed. So transfection duration was estimated based on cells growth rate and knock down efficiency for various genes. Experiments were carried out beforehand for each gene to test at what concentration of siRNA knock down was optimal.

3.3. Western blot

For estimation of protein expression western blot was used. Protein expression was estimated both for correlation tests and for measurements of knock down efficiency. Correlation tests that were carried out for the LETM2 and the PPAPDC1B genes were either done on western blot membranes previously prepared by Berglind Ósk Einarsdóttir (16) as well as new western blot membranes that were made using lysates extracted by Berglind Ósk Einarsdóttir in connection with her MSc. project (16). For estimation of Ppapdc1B expression in normal tissue, protein extraction was carried out by the author of this thesis, Edda Olgudóttir. For estimation of protein knock down protein extraction from cell lines was done by Edda Olgudóttir. As stated previously protein extraction from MEP, LEP and D492-3D was done by the team of SCRU at Læknagarður.

3.3.1. Protein extraction

For protein isolation RIPA buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1.0% Nondet P-40 (NP40), 1.0% Na-deoxycholate and 0.1% sodium dodecyl sulfate (SDS)) was used. 7x protease inhibitor (PI) cocktail (Roche Applied Sciences) was diluted to the RIPA buffer as well as 100x dithiothreitol (DTT) (Sigma-Aldrich). The PI was added to prevent protein degradation and the DTT was used to prevent formation of disulfide bonds.

3.3.1.1. Cell lines

Cell culture medium was removed and the cells were washed twice with cold PBS before put on ice. The RIPA buffer plus PI and DTT was added to each well, 10 µl for 96-well plate and 50 µl for 24-well plate or into T-25 cell culture flasks, 500 µl in each flask. Cell culture plates or flasks were gently rotated in order to cover all cells with RIPA buffer. The samples then stayed on ice for approximately 10 minutes, before being transferred to eppendorf-glasses and kept on ice for approximately 20 minutes. In order to break down DNA molecules the lysates were squeezed through a 25 G needle several times. After that the lysates were centrifuged for 40 minutes in 4°C, 14.000 rounds per minutes (rpm). Then the supernatant was collected and the pellet discarded. Lysates were stored in -80°C.

3.3.1.2. Tissue samples

Tissue samples were collected from breast cancer patients diagnosed in the years 1991-2000. The samples were collected at the Department of Pathology at Landspítali University Hospital. The

samples were kept in aluminum bags and were kept frozen in -80°C before protein extraction was carried out. For protein isolation a part of the tissue was cut off using a scalpel and put on a Petri dish which was placed on dry ice to keep the samples frozen. The samples were then transferred to pre-cooled beakers that had been dipped in liquid nitrogen (LN_2). The beakers were then again dipped in LN_2, containing the samples, before placed in the Micro Dismembrator S (Sartorius) for 30 seconds at 2600 rpm. After the beakers had been cooled again in LN_2 they were put in the Micro Dismembrator for another 30 seconds. The tumor sample was then dissolved in 300-400 µl RIPA buffer, containing both PI and DTT. The lysate was then treated as described previously for protein extraction in cell lines.

3.3.1.3. Protein concentrations estimation

Protein concentrations were measured with bicinchoninic acid (BCA) protein assay. For this Pierce BCA Protein Assay Kit (Thermo Scientific) was used. The BCA protein assay is based on the reduction of Cu^{+2} to Cu^{+} in the presence of a protein in an alkaline solution. The protein lysates are mixed with a solution containing Cu^{2+} ions and bicinchoninic acid. The proteins catalyze the reduction of Cu^{2+}, the Cu^{+} ion that forms can then bind two BCA molecules. The Cu^{+}BCA complex exhibits absorbance at 562 nm, which can be measured in Nano Drop 1000 spectrophotometer. To minimize the lysates used they were first diluted ten times in RIPA buffer containing PI and DTT. The dilution was then added to the working reagent, mixed according to the manufacture's protocol. This mixture was then incubated at 37°C for 30 minutes. The samples were cooled to room temperature before the protein concentrations were measured in Nano Drop. The same procedure was carried for series of five Bovine Serum Albumin (BSA) standards, solutions that contain BSA of known concentrations for the making of standard curves.

3.3.1.4. Preparation of samples

Lysates were diluted with RIPA buffer to yield 1 µg/µl mixture before electrophoresis. 1x Protein sample buffer (PSB) (60 mM Tris (pH 6.8), 2% SDS, 10% glycerol, 0.008% bromophenol blue, 5% β-mercaptoethanol) was added to the samples. The PSB buffer plays several roles during the electrophoresis, 5% β-mercaptoethanol breaks disulfide bridges.The SDS which is negatively charged binds to the proteins giving them negative charges. Then the proportion of charge versus weight is the same for all proteins. This negative proportion of charge and weight results in proteins of different size running at different velocity through the gel, see below. In addition the glycerol is denser than the surrounding buffer which makes it easier to load the samples on the gel and bromophenol blue helps visualize the samples when loading. The pH value for the PSB is the same as the pH value for the stacking gel, see below.

3.3.2. Gel preparation

SDS-PAGE (Sodium-dodecyl-sulfate Poly-acrylamide gel electrophoresis) was used to separate proteins according to size. Different concentrations (10-15%) of running gels were used for each protein depending on the size of the target protein. The concentration depends on the amount of acrylamide compared to the final volume. Below is an example of a gel recipe for 10% gel, for different

concentrations changes should be made on the acrylamide and water volume. The stacker gel (5%) was always used as described in Table 2

Table 2. A recipe for 10% running SDS- PAGE and 5% stacking SDS-PAGE gel.

	10% SDS-PAGE	**Stacking SDS-PAGE**
H_2O	2.0 ml	0.68 ml
30% Acrylamide mix*	1.7 ml	0.17 ml
1.5 M Tris (pH 8.8)	1.3 ml	
1.0 M Tris (pH 6.8)		0.13 ml
10% SDS	0.05 ml	0.01 ml
10% APS**	0.05 ml	0.01 ml
TEMED***	0.002 ml	0.001 ml
Final volume	5 ml	1 ml

* 29.2% acrylamide and 0.8% N,N'-methylene-bis-acrylamide;
** Amonium persulfate;
*** N,N,N',N'-Tetramethylethylenediamine.

For making of the gel the first four ingredients were mixed and the solution was degassed for 15 minutes at 60 kPa before both APS and TEMED were added. The solution then polymerized, for 45 minutes, between two glass plates to make 0.75 mm thick gels. For the separation of Ppapdc1B a gradient gel was used (4-15%) which was purchased from Bio-Rad.

The SDS in the gel, like in the PSB, binds to denatured proteins. The proteins then all carried the same net charge in proportion to their molecular weight.

3.3.3. Electrophoresis and transfer

5 µg of protein samples were loaded per lane on the SDS-PAGE gel along with Page Ruler pre-stained protein marker (Fermentas) for size estimation. Ppapdc1B and Ash2l detections were an exception but in that case 10 µg of protein were used. For electrophoresis Mini PROTEAN Tetra Cell from Bio-Rad were used. The gel was placed in the cell which was filled with running buffer (25 mM Tris base, 192 mM Glycine, 1 % SDS). After the loading of samples proteins were run through the stacking gel on 100 Volts for approximately 20 minutes. The voltage was then raised up to 150 Volts and the lysates ran through the running gel for approximately one hour, depending on the gels density. When proteins traveled through the acrylamide complex their velocity depended on the proteins molecular weight. The density of each gel contributed to this velocity by increasing the difference between proteins of different molecular weight.

After electrophoresis the proteins were transferred onto a Polyvinylidene fluoride (PVDF) Immobilon-P membrane (Millipore), with 0.45 µm pore size. The SDS-PAGE gel was first wet in transfer buffer (25 mM Tris, 192 mM glycine). The PVDF membrane is water resistance and was therefore first put in methanol for 15 seconds then transferred to water for two minutes and at last wet

in transfer buffer. For transfer of proteins the PVDF membrane and SDS-PAGE gel were packed between six chromatography papers (Whatman), three sheets on each side, all previously wet in transfer buffer. The transfer was performed in TRANS BLOT (BioRad), at 15 Volts for 40-60 minutes, depending on the size of the target protein.

3.3.4. Antibody probing

After transfer, the membrane was incubated, in Tris buffered saline with added Tween-20 (TBS-T) (250 mM Tris (pH 8.0), 1.5 mM NaCl, 0.1% Tween-20) in which 5% non fat milk powder was added, for one hour for membrane blocking. This was done to prevent non-specific binding of the antibodies to the PVDF membrane. The membranes were then probed with primary antibodies overnight; antibodies used are listed in

Table 3. The protein concentration varied between each antibody, see Table 3. After primary probing the membranes were washed four times for five to ten minutes each time. Then probing with Horseradish Peroxidase labeled secondary antibody was carried out for approximately one hour. Secondary antibody concentrations are listed in Table 3. All antibody incubations were carried out at 10-15°C and on an orbital shaker. Washing of the membrane, after incubation of the secondary antibody was then carried out as described before. All antibody incubations were done in 5% non fat milk in TBS-T, in order to minimize background signals.

All antibodies were tested previously in order to determine the concentration for incubation. The antibody for Ppapdc1B was tested on extracts from cerebral cortex from rat, which was a gift from the laboratory of Pétur Henry Petersen, at the Neurobiology at Læknagarður.

Table 3. A list of the antibodies used for this study.

	Concentration for incubation	Raised in	Purchased from
Primary antibodies			
α-Letm2	0.05 µg/ml	Rabbit	Atlas Antibodies
α-Ppapdc1B	1.00 µg/ml	Rabbit	Aviva Systems biology
α-Ppm1d	0.07 µg/ml	Rabbit	Novus Biologicals
α-Plk1	0.03 µg/ml	Rabbit	Santa Cruz Biotechnology Inc.
α-Erlin2	0.5 µg/ml	Goat	Abcam
α-LSM1	1.00 µg/ml	Rabbit	Stragetic diagnostic Inc
α-Ash2l	2.00 µg/ml	Mouse	Abcam
α-PROSC	0.33 µg/ml	Rabbit	Stragetic diagnostic Inc
α-actin (C4 clone)	0.034 µg/ml	Mouse	ICN Biomedicals
α-actin (C4 clone)	0.5 µg/ml	Mouse	Abcam
α-Gapdh	0.057 µg/ml	Mouse	RDI division of Fitzgerald industries Intl.
Secondary antibodies			
α-mouse	0.075 µg/ml	Goat	Dako Cytomation
α-mouse	0.075 µg/ml	Rabbit	Invitrogen
α-goat	0.075 µg/ml	Rabbit	Invitrogen
α-rabbit	0.075 µg/ml	Goat	Invitrogen

For detection of the proteins the membranes were incubated for one minute in ECL plus (GE Healthcare) solution. When the membranes were exposed to the ECL solution the following reaction took place: The ECL solution contained Lumigen PS-3 Acridan which underwent oxidation in the presence of HRP and peroxide. This reaction generated CO_2 and light (Chemiluminascence). The membrane was then packed in plastic wrap and placed in x-ray film cassette. The following steps were performed in a dark room. Each membrane was exposed to a photographic film (GE Healthcare) for various exposure times in order to visual both weak and strong signals. The films were then run through a film processor (DURR MED 260) where they were developed.

3.3.5. Signal quantification

The films were scanned in Canon Scan LiDE 90 (Canon) on a TIFF format, in 600 dpi resolution. They were opened in Photoshop CS2 (Adobe) where the photos were put on grayscale and inverted. The inverted bands were shown as light, thus more protein gave more light. A circle was manually drawn

around the signals, Photoshop then gave both pixel count within each band and a mean value which represents the intensity of the signal, (higher intensity equals more protein). This information was used to calculate relative protein expression. For each signal, background was measured as well by using the same methods as for the signals, the same pixel amount was used but the mean represented the signal around the protein bands. All measurements were collected into Excel (Microsoft) where calculations were done.

To calculate the intensity of the protein signal:

[pixels] * [mean (protein signal)]

To calculate the intensity of the background:

[pixels] * [mean (background signal)]

To calculate the relative protein signal:

[protein signal] – [background signal]

These calculations were done for the target proteins as well as the controls, β-Actin and Gapdh (housekeeping proteins). The expression of target proteins was standardized to the expression of the controls. In order to standardize the protein expression the target proteins' signals were used as a portion of β-Actin/Gapdh expression in each sample:

[target protein signal] / [loading control signal]

This was done to avoid bias due to uneven loading on SDS-PAGE or correct for bias in protein concentration measures. The loading control used in this project was β-Actin with one exception where Gapdh was used, when PPM1D was downregulated.

For correlation tests of Ppapdc1B and Letm2 three to four membranes were used to measure the protein expression in 21-36 tumor samples. In this case one cell line (SUM52) was loaded on each gel on comparison of the protein expression between gels. To do so the protein expression in SUM52 was set as one. For measurements of protein expression in cell lines used for siRNA knock down the D492-3D cell line was used for comparing protein expression between cell lines and set as one. For downregulation tests, the protein expression in cells treated with scrambled siRNA was set as one.

3.4. Correlation calculations

For Ppapdc1B and Letm2 the protein expression data were used to analyze correlation between DNA status or mRNA expression and protein expression in the breast tumor samples. Information on gene copy number and relative amount of mRNA was collected from an online database BioArray Software Environment (BASE) (http://base.onk.lu.se/onk/). In this database aCGH and aGEx results from analysis of more than 500 tumor samples are kept. These tumor samples were from our lab as well as

other labs from Sweden and Finland (15). This database is held by the laboratory of Åke Borg Department of Oncology, Clinical Science, at Lund University.

The data was collected into Excel where DNA, mRNA and protein data for each sample was matched. Samples classification based on DNA data went as follows: Loss = DNA ratio < 0.8, Normal DNA levels = DNA ratio 0.8 – 1.2, Gain of DNA levels = DNA ratio 1.2 – 1.8 and Amplification = DNA ratio > 1.8. Data were transformed to Log2 scale in Excel for correlations calculations. Correlation test was carried out using Pearsons' product moment correlation coefficient (r) function on Excel. This was done for DNA and protein values for both genes and also mRNA values for Ppapdc1B. The p-value calculations were done on a free online calculator Statistics calculator's version 3.0 beta (http://danielsoper.com/statcalc3/default.aspx). Correlation was estimated statistical significant if the p-value was below 0.05.

3.5. Cell proliferation test

At first the cell proliferation tests were carried out using MTT-assay (Thiasolyl Blue Tetrazolium Bromide), but results were always verified with cell counting. Cell counting was then used as the primary method for measuring changes in proliferation later on. This was done in order to be able to expand the downregulation time when needed.

Cell proliferation after downregulation of ERLIN2 and LSM1 were measured with MTT-assay. However when ASH2L, PROSC, and ALP (ASH2L, PROSC and LSM1), were downregulated the cell proliferation was measured with cell counting only. In addition effects of the PPM1D knock down in MCF7 were measured with MTT-assay but by counting in ZR-75-1. For PLK1 downregulation MTT-assay was used to estimate the effects in MCF7, CAMA-1 and ZR-75-1 but cell counting was used for SUM52 and T47D.

3.5.1. MTT-assay

After transfection the cells are cultured in suitable medium without phenol red (Gibco-Invitrogen) containing the right supplements for each cell line in 96-well cell culture plates. When knock down had been held out for appropriate time, 100 µg MTT, dissolved in PBS was added to each well and incubated for 3.5 hours. When the MTT solution was incubated a reduction of the MTT occurs. Before the MTT is reduced the color of the solution is yellow. When MTT solution is added to the cells, active mitochondrial reductases reduce the molecule, yielding a purple formazan. The media was then removed and 100 µl MTT solvent (4 mM HCl, 0.1% NP40, dissolved in isopropanol) was added to each well. When the cells were lysed in the MTT solvent the purple formazan is released from the cells yielding a purple solution. The intensity of the purple color was relative to the amount of cells with active reductase, in other words cells that were alive. The cell culture plates were agitated for approximately 15 minutes, wrapped in tinfoil to protect the solution from light exposure. Cells and MTT-solvent were mixed vigorously with pipette if needed. The solutions absorbance was then read at 590 and 620 nm in Spectra Max Plus 384 (Molecular Devices). For all treatments one negative control

(neg-control) well was used, containing only MTT solution and solvent but no cells, this well represented the background.

After the absorbance was read all data were collected in Excel where the following calculations were done:

[590 nm value]-[neg-control 590 nm value]-[620 nm value]-[neg-control 620 nm value]

The average value was found for three wells treated the same way. Standard deviations were calculated from the same three wells. Cells treated with scrambled siRNA were set as one.

3.5.2. Cell counting

After transfection the cell proliferation was estimated by dispensing the cells in volumes varying from 200-500 µl. The cell solution was mixed gently for few a seconds before a portion was spread on a hemocytometer slide. A hemocytometer is a slide which has been cut in order to divide the surface area of 9 mm^2 into nine 1 mm^2 squares. Each of these squares is also divided into 25 equal squares for easier counting. The 1 mm^2 chambers were used to count the cells. When a sample is put on the hemocytometer a coverslip is placed over the slide, making a chamber for the cell volume. The depth of the chamber is 0.1 mm, so the volume in which is counted each time is:

$1\ mm^2 * 0.1\ mm = 1\ mm^3$ or $1\ mm^2 * 0.1\ mm = 1 * 10^{-4}$ ml

The cells were viewed in Leica DM IL LED inverted microscope and counted. Samples from each well were counted at least three times. Cells were re-mixed each time a sample was taken for counting. In order to estimate the cell number in each well:

Cells /ml = [Cell count on 1 mm^2 hemocytometer] * 10^4

The cells were treated in triplicate, when knock down was performed. For each treatment the average value was calculated based on three wells and standard deviations were also found. The cell proliferation of cells treated with scrambled siRNA was set as one. All calculations were carried out in Excel.

4 Results

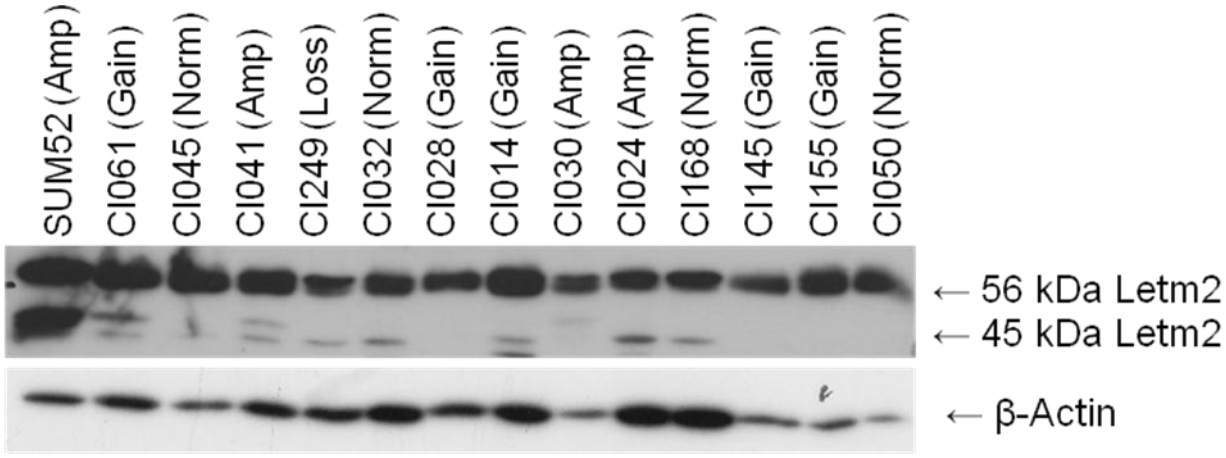
Previous studies in our lab narrowed the amplification on chromosome 8p12-p11 to a region containing 20 genes. Eighteen of them had been analyzed further and of those seven were identified as putative target genes of the amplification. Two genes within the locus remained to be analyzed with regards to protein expression, LETM2 and PPAPDC1B, and thus could be potential targets as well. Therefore, the search for the amplification´s target gene continued with protein expression analysis of these two genes prior to the analysis of oncogenic potential of the putative oncogenes in breast cancer cell lines.

4.1. PPAPDC1B and LETM2

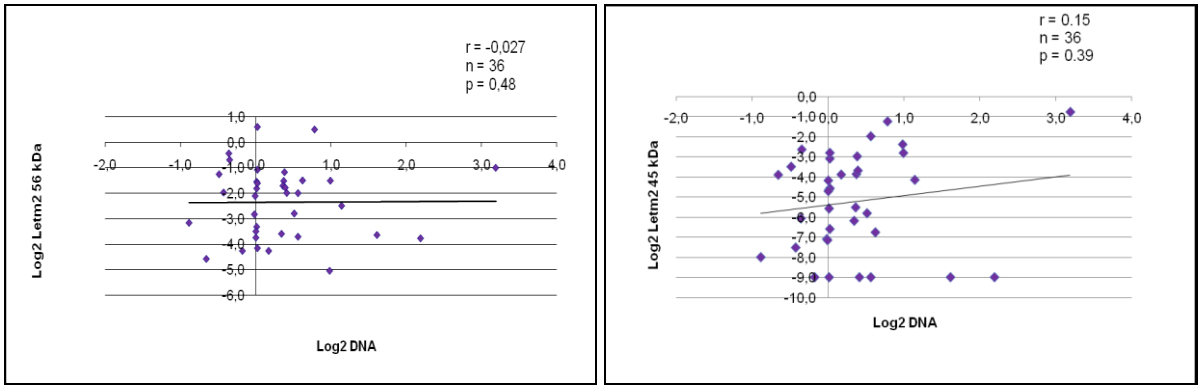
Two of the 20 genes that are located within the amplified 8p12-p11 region are the genes LETM2 and PPAPDC1B. When the protein studies by Einarsdottir (2010) were performed no commercial antibodies against Letm2 and Ppapdc1B were available. Antibodies became available recently and it was tested whether these genes showed correlation between DNA amplification, mRNA quantity and protein expression.

4.1.1 LETM2 is not a target gene of the 8p12-p11 amplification

aGEx mRNA data for LETM2 on BASE were only of sufficient quality in a few samples and therefore mRNA could not be correlated to DNA amplifications. To test if LETM2 could be the 8p12-p11 amplicon´s target gene we studied its protein expression to correlate it to the DNA levels. We used 36 breast tumors to analyze the Letm2 protein expression on western blot (Figure 1). DNA from all 36 tumors had undergone aCGH and data were obtained from BASE, to analyze the correlation between DNA status and protein expression. The classification of loss, normal, gain and amplification of DNA copy number is described in materials and methods.



a)



b) c)

Figure 1. Letm2 protein levels are not elevated when LETM2 gene is amplified at the DNA level in breast tumors. Protein lysates (5 µg/lane) from 36 breast tumors were loaded on SDS-PAGE for electrophoresis, the proteins were transferred to PVDF membrane which was probed using antibodies against Letm2 (0.05 µg/ml) and β-Actin (0.034 µg/mL). Protein levels were quantified as described in materials and methods. a) An example of the western blots that were used to analyze the Letm2 and β- Actin protein expression, showing 11 breast tumors and one cell line, SUM52. DNA status, for LETM2. is indicated in brackets after each samples name, Loss = loss of copy number, Norm = normal copy number, Gain = gain of copy number and Amp = DNA amplification. b) Correlation analysis of LETM2 DNA status versus protein expression of the 56 kDa isoform (r = -0.027 (p=0.48)) c) and the 45 kDa isoform (r = 0.15 (p = 0.39)). Protein expression was standardized using β-Actin as loading control. DNA results in each sample are aCGH results obtained from BASE. The results are displayed on Log2 scale. Pearson's correlation coefficient was used to estimate correlation, r. n = the number of samples analyzed.

When we performed these experiments only two transcripts were known for the LETM2 gene, coding for proteins of 56 kDa and 45 kDa molecular weight. Based on this information, experimental tools were used that could only detect proteins above 40 kDa molecular weight. Information about the Letm2 protein that have been published since our experiments were performed, suggest that the LETM2 gene codes for nine different isoforms (http://sep2011.archive.ensembl.org/index.html) (101). These different isoforms vary from 13 kDa to 56 kDa. According to the producer of the antibody it was supposed to detect a protein of approximately 56 kDa molecular weight. When we first tested the antibody this 56 kDa protein was exactly what we observed, as well as a weak signal just below the 45 kDa marker. This 45 kDa signal agrees with published data on LETM2 (101). However this signal was not observed in all of the samples used (Figure 1 and 2). The 45 kDa was so weak, in the breast

tumors, compared to the signal given by SUM52 that most of the values are below zero when transformed to Log2 scale.

When western blot was performed on breast tumor samples both signals were quantified. No correlation between the LETM2 gene copy number and 45 kDa protein levels was observed, in the 36 breast cancer samples that were used for western blotting, (Figure 1). This was also the case when the 56 kDa isoform was quantified and correlation calculations were done on gene copy number and protein expression (Figure 1).

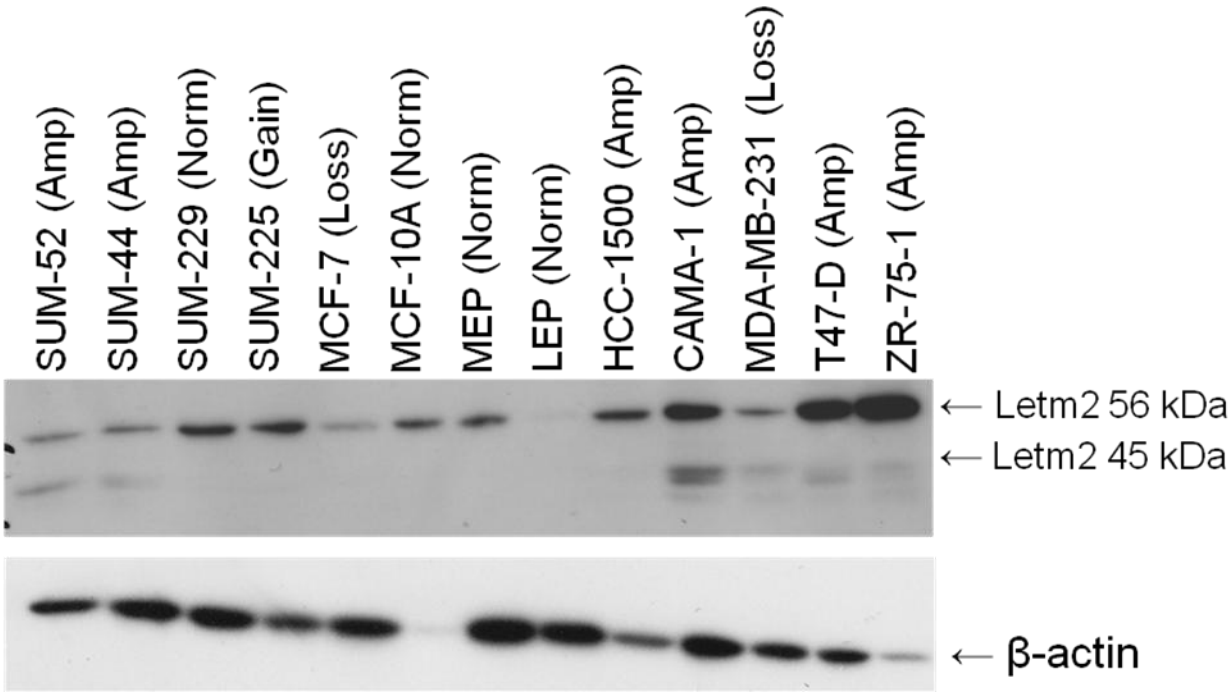


Figure 2. Letm2 protein expression in breast cancer cell lines. Protein lysates (5 µg/lane) from 13 cell lines, derived from healthy or cancerous breast tissue, were loaded on SDS-PAGE for electrophoresis and then transferred to PVDF membrane that was blotted with antibodies against Letm2 (0.05 µg/ml), above, and β-Actin (0.034 µg/mL), below. Protein expression was estimated as described in materials and methods. DNA status, for LETM2 is indicated in brackets, behind cell names, Loss = loss of copy number, Norm = normal copy number, Gain = gain of copy number and Amp = DNA amplification.

All the breast cancer cell lines analyzed displayed the 56 kDa Letm2 isoform but no correlation between the DNA gene copy number and protein expression was observed. All cell lines carrying amplification of the LETM2 gene displayed signal from the 45 kDa Letm2 isoform, with the exception of HCC1500 (Figure 2). Two cell lines that should according to our hypothesis be expressing more LETM2, HCC1500 (amp) and SUM225 (gain), did not display any signal at 45 kDa. This was also the case for cell lines that carried normal copy number and a loss for this gene, with the exception of MDA-MB-231 (Figure 2). The expression of the 56 kDa protein in the cell lines support the results observed in the breast tumors. The same is true for the 45 kDa protein even though the protein expression results from the cell lines were not as convincing.

Based on these results the 56 kDa and 45 kDa isoforms of Letm2 were excluded as the target of amplification on chromosomal region 8p12-p11.

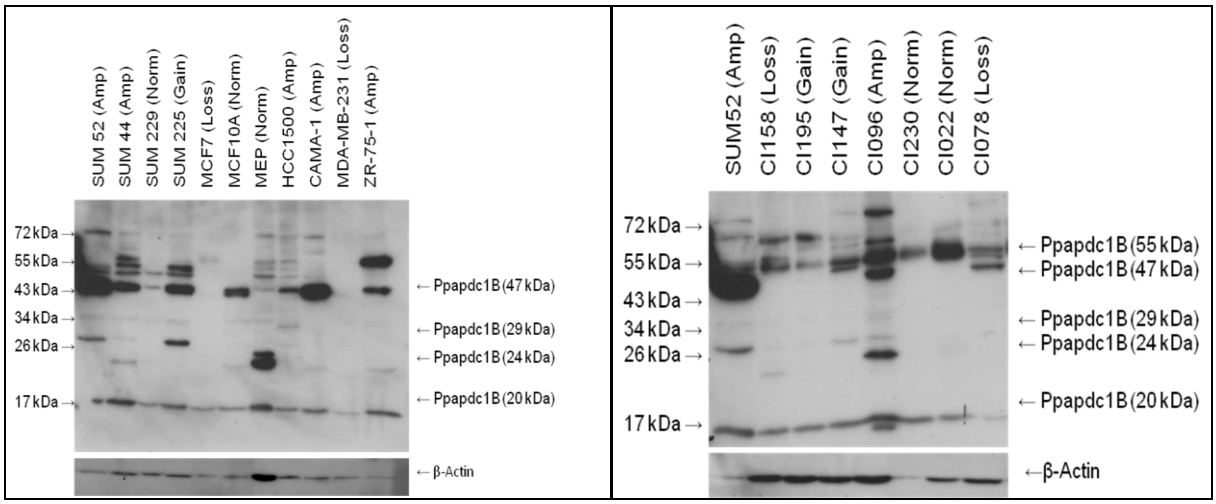
4.1.2. PPAPDC1B is more likely a tumor suppressor gene than an oncogene

As described in the introduction chapter, some groups have pointed the PPAPDC1B gene out as a target gene for 8p12-p11 amplification in breast tumors (39). However, the gene has also been implicated as a metastatic tumor suppressor gene in HCC (56, 57). Analysis done in our lab on 161 breast tumor samples showed significant correlation between PPAPDC1B DNA and mRNA levels ($r = 0.64$, $p < 0.001$) (16). To find out if this correlation also existed at the protein level we analyzed the Ppapdc1B expression in 36 of these 161 breast tumors using western blot.

4.1.2.1. PPAPDC1B codes for few different isoforms

According to the ensembl database PPAPDC1B codes for seven different isoforms that range from 15 kDa to 29 kDa (101). Only three isoforms however have been anticipated based on the gene structure and therefore received a Consensus coding sequences (CCDS) number, that weigh approximately 29 kDa, 24 kDa and 20 kDa (http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi Sept. 2011) (102). It should be noted that even though isoforms have not received a CCDS number it is possible that they will gain one later when more studies have been done. We used a commercial antibody targeting Ppapdc1B. The antibody was raised against an epitope that all three accepted isoforms have in common, as well as additional isoforms listed on the ensembl database, with the exception of the 15 kDa isoform (101). When we analyzed the Ppapdc1B expression in the breast tumors we observed an unexpected western blot band pattern that did not fit the isoforms listed in databases (Figure 3). The strongest signal detected by the antibody was around 47 kDa, in breast cancer cell lines (Figure 3a) but approximately 55 kDa in breast tumors (Figure 3b).

The signals at 55 kDa and 47 kDa were unexpected so to verify that the antibody we were using did detect Ppapdc1B we first used BLAST to see if the immunogen used to generate the antibody could match any other protein that might give a signal at 55 kDa on western blot. No protein matched the immunogen other than Ppapdc1B and Ppapdc1A, which showed 40 similar amino acids out of 50 possible. The 10 amino acids missing were located between other amino acids in the middle of the immunogen so it is unlikely that the antibody was detecting any other protein than Ppapdc1B, although not excluded, based on these results alone. In addition Ppapdc1B is approximately the same size as Ppapdc1B, or 30 kDa so the signal we got at 55 kDa was unlikely to be Ppapdc1A (101).



a) b)

Figure 3. Ppapdc1B protein expression in breast cancer cell lines and breast tumors is variable. Protein lysates (10 µg/lane) from eleven breast cancer cell lines (a) and seven breast tumors (b) were loaded on SDS-PAGE for electrophoresis and transferred to PVDF membrane that was blotted with antibodies against Ppapdc1B (1.0 µg/ml) shown above and β-Actin (0.5 µg/ml) shown below. Samples names are listed above western blots, DNA status is indicated in brackets, Loss = loss of copy number, Norm = normal copy number, Gain = gain of copy number and Amp = DNA amplification. Protein marker size is indicated on the left and on the right Ppapdc1B isoforms are indicated.

Ppapdc1B is a well conserved protein, for example Ppapdc1B in humans, mice and rats are highly similar. Rat Ppapdc1B is, like the human protein, supposed to be smaller than 30 kDa (101). When we tested the rat tissue on western blot we found that the antibody detected a protein at the right size, approximately 24 kDa (**Figure 4**). On the same gel a strong 47 kDa band was seen from ZR-75-1 (**Figure 4**). These observations supported that our antibody is detecting Ppapdc1B in human tissue, even though the signal was not of the right size.

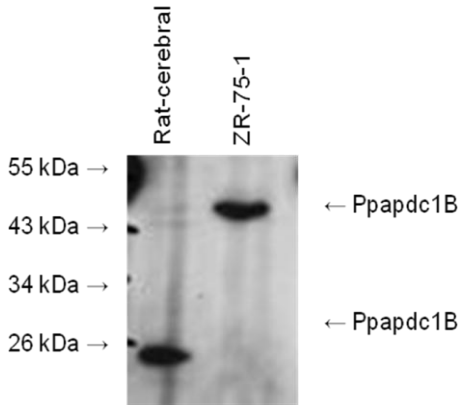


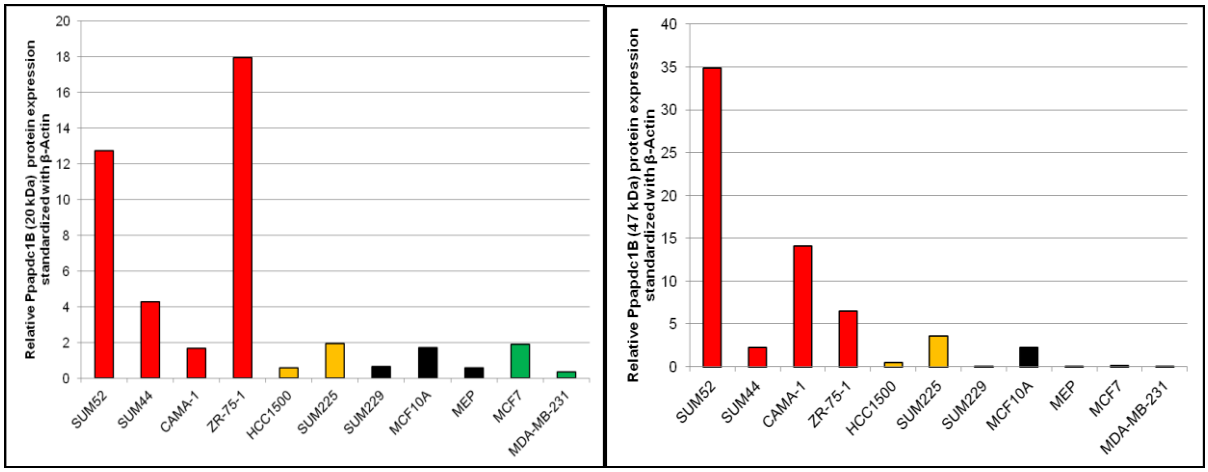
Figure 4. Antibody against Ppapdc1B detected right sized isoform in rat tissue as well as bigger isoform in breast cancer cell line. Protein lysates (10 µg/lane) from rat cerebral cortex and breast cancer cell line, ZR-75-1 were loaded on SDS-PAGE for electrophoresis and then transferred to PVDF membrane that was blotted with antibody against Ppapdc1B (1.0 µg/ml).

To confirm our suspicion that these large sized bands were isoforms of Ppapdc1B we also tested another antibody targeting Ppapdc1B that had been raised against a different epitope. This antibody also targeted all the same isoforms. We got similar result with both antibodies, i.e. bands of 47 and 55 kDa molecular weight (data not shown). Based on these observations we concluded that

the strongest signal observed, at 55 kDa, for breast cancer samples or 47 kDa for breast cancer cell lines, was most likely Ppapdc1B.

4.1.2.2. Ppapdc1B protein expression correlates with amplification in breast cancer cell lines

The Ppapdc1B expression patterns in breast cancer cell lines varied from one cell line to another (Figure 3a). The 20 kDa isoform was displayed in all cell lines (signal observed around 17 kDa marker in Figure 3). After quantification we found that cell lines carrying amplification of chromosomal region 8p12-p11 expressed higher levels of Ppapdc1B – 20 kDa than cells that carry gain, normal or loss of DNA levels with the exception of CAMA-1, (Figure 5a).



a) b)

Figure 5. The expression of Ppapdc1B, in breast cancer cell lines correlates with DNA status. Protein lysates (10 µg/lane) from eleven breast cancer cell lines were loaded on SDS-PAGE for electrophoresis and transferred on to PVDF membrane which was blotted with antibodies against Ppapdc1B (1.0 µg/ml) and β-Actin (0.5 µg/ml). Signals at both 20 kDa (a) and at 47 kDa (b) were quantified as described in materials and methods and standardized with β-Actin. Cell lines are ranked according to their DNA status on PPAPDC1B. Red columns show cells harboring amplification, yellow shows cells carrying a gain on PPAPDC1B, black represents normal DNA levels and green denotes a loss (Table 5).

The 24 kDa isoform was highly expressed in MEP, while only weak signals could be observed at this size for the breast cancer cell lines SUM44, CAMA-1 and ZR-75-1, and the normal cell line MCF10A. Other cells did not display a signal for 24 kDa (Figure 3a). Finally, only two cell lines, SUM52 and SUM44, showed signals that might be interpreted as the 29 kDa isoform (Figure 3a). Statistical analysis was not done for those isoforms because they were only expressed in a few cell lines.

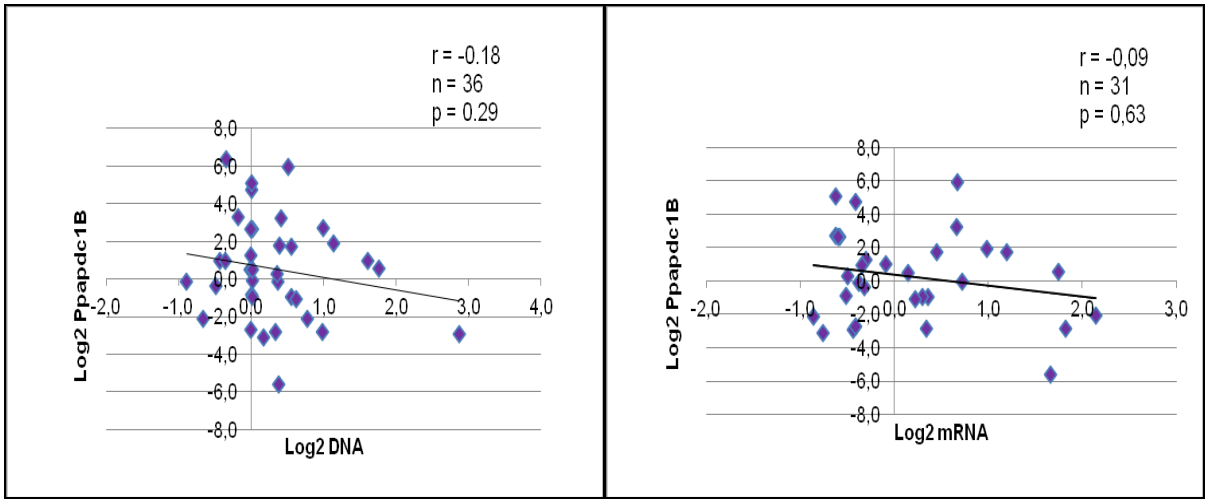
All cell lines expressed a protein that gave a signal at approximately 47 kDa except for MEP and two cell lines that carry a loss on chromosomal region 8p12-p11 (MCF7 and MDA-MB-231), (Figure 3a). The two cell lines with normal copy number of PPAPDC1B, SUM229 and MCF10A gave either a weak signal at 47 kDa (SUM229) or expressed high levels of the 47 kDa band (MCF10A) (Figure 3a). All cell lines carrying amplification expressed high levels of 47 kDa Ppapdc1B, except SUM44 (Figure 5b). However HCC1500 that carries a gain on PPAPDC1B did not show high levels of the 47 kDa band (Figure 5b).

One of the conditions set for a putative target gene in this study is that protein expression in cell lines harboring amplification had to exceed protein expression in cell lines that carried loss or normal copy number on this region. This had to be observed in at least three out of four cell lines carrying amplification (16). When studying Ppapdc1B both 20 kDa and 47 kDa isoforms fulfilled this condition (Figure 5). We therefore included PPAPDC1B as a possible target gene, based on the correlation between DNA status and Ppapdc1B protein expression in breast cancer cell lines.

4.1.2.3. Ppapdc1B protein expression in human breast tissue

For the breast tumors, variable expression patterns of Ppapdc1B protein expression were also observed (Figure 3b). As mentioned above, breast tumors displayed a relatively strong signal at approximately 55 kDa molecular weight which was observed in all tumor samples (Figure 3b). We excluded the possibility that this signal was background signal from the secondary antibody, like IgG, by blotting separately with secondary antibody only. This blot did not display any signal, the 55 kDa signal was exclusive to the primary antibody. Figure 3b shows an example of different protein expression patterns in the breast tumor samples. In all seven tumor samples the 20 kDa isoform was expressed (Figure 3b). Gene copy number data, taken from BASE, for this particular gene were available for all breast tumors used for western blotting. When the western blot was performed to estimate the abundance of the 20 kDa isoform, protein lysates were available for 21 tumor samples. The protein levels in these tumors did not display correlation with the DNA quantity ($p = 0.33$). The mRNA data, obtained from BASE were available for 17 tumors of these 21 tumor samples. No correlation was observed between mRNA and protein levels ($p = 0.21$). The aGEx mRNA probe used should detect all known transcripts from the PPAPDC1B gene. Other accepted isoforms were not expressed widely enough to do statistical analysis.

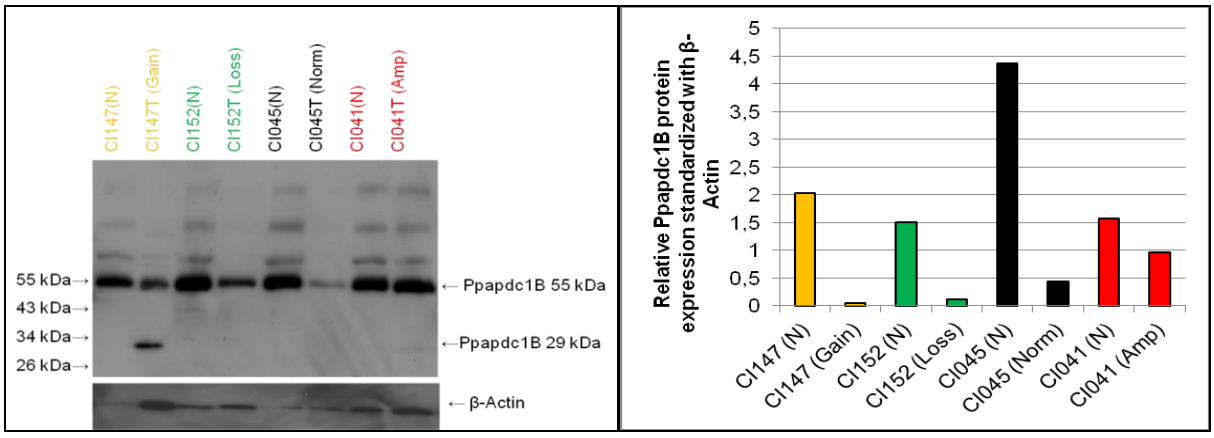
The 55 kDa signal was quantified to do correlation calculations on PPAPDC1B in breast tumors. Data from 36 tumor samples were available to do protein measurements. Figure 6a shows that no correlation was detected between protein expression and the DNA status in breast tumor samples. This was also true for correlation between protein and mRNA expression as shown in Figure 6b. mRNA data were available for 31 tumor sample.



a) b)

Figure 6. PPAPDC1B protein expression (55kDa) does not correlate with DNA levels on chromosomal region 8p12-p11, or mRNA expression, in breast tumors. Protein lysates (10 µg/lane) from 36 breast tumors were loaded on SDS-PAGE and transferred to PVDF membrane that was blotted with antibodies against Ppapdc1B (1.0 µg/ml) and β-Actin (0.5 µg/ml). Expression of the 55 kDa Ppapdc1B was estimated as described in materials and methods and standardized with β-Actin. DNA and mRNA data for 36 and 31 tumors, respectively were obtained from BASE. Results are displayed on Log2 scale. Correlation was calculated using Pearson's coefficient. a) Ppapdc1B protein expression and DNA status in 36 breast tumos. Pearson's correlation coefficient r = -0.18, (p = 0.29). b) Ppapdc1B protein and mRNA expression in 31 breast tumors Pearson's correlation coefficient, r = -0.09 (p = 0.63). n = the number of samples analyzed.

Even though mRNA levels are elevated when PPAPDC1B is amplified the levels of the 55 or 20 kDa isoforms of Ppapdc1B are not increased. Ppapdc1B had awoken our interest based on the correlation the protein expression showed in breast cancer cell lines. We therefore decided to compare Ppapdc1B protein amounts in normal breast tissue and breast tumors (Figure 7).



a) b)

Figure 7. Ppapdc1B, 55 kDa, is less expressed in breast tumor samples than their matching normal breast samples. Protein lysates from four breast tumors and their matching normal tissue (marked N in brackets) were loaded on SDS-PAGE (10 µg/lane) for electrophoresis and transferred to PVDF membrane that was blotted with antibodies against Ppapdc1B (1.0 µg/ml) and β-Actin (0.5 µg/ml). Colors indicate DNA status in each tumor sample, yellow represents samples with gain, green samples with loss, black is for normal DNA status and red means amplification. The matching normal sample (N) is given the same color as tumor samples. a) Western blot of breast tumors and their matching normal tissue. DNA status in each case is indicated in brackets, DNA status in normal samples is assumed to be normal. b) Protein levels of 55 kDa Ppapdc1B were quantified and standardized with β-Actin.

The 55 kDa isoform of Ppapdc1B was highly expressed in normal tissue when compared to matching tumor samples (Figure 7). These observations are in agreement with published data from a Chinese group that speculated that PPAPDC1B was a metastatic tumor suppressor gene (56, 57). We verified these results in a normal tissue that was not taken from breast cancer patients as was the case for the samples in Figure 7. Protein was isolated from breast tissue taken from reduction mammoplasty and Ppapdc1B abundance was tested with western blot. Normal tissue showed relatively high levels of Ppapdc1B, 55 kDa, like we observed in the normal samples taken adjacent to tumors (data not shown).

Based on the results from our analyses PPAPDC1B might be a tumor suppressor protein in breast tumors. We did not observe elevated protein levels in breast tumors carrying amplification of the PPAPDC1B gene. We therefore excluded PPAPDC1B as a 8p12-p11 amplicon target gene.

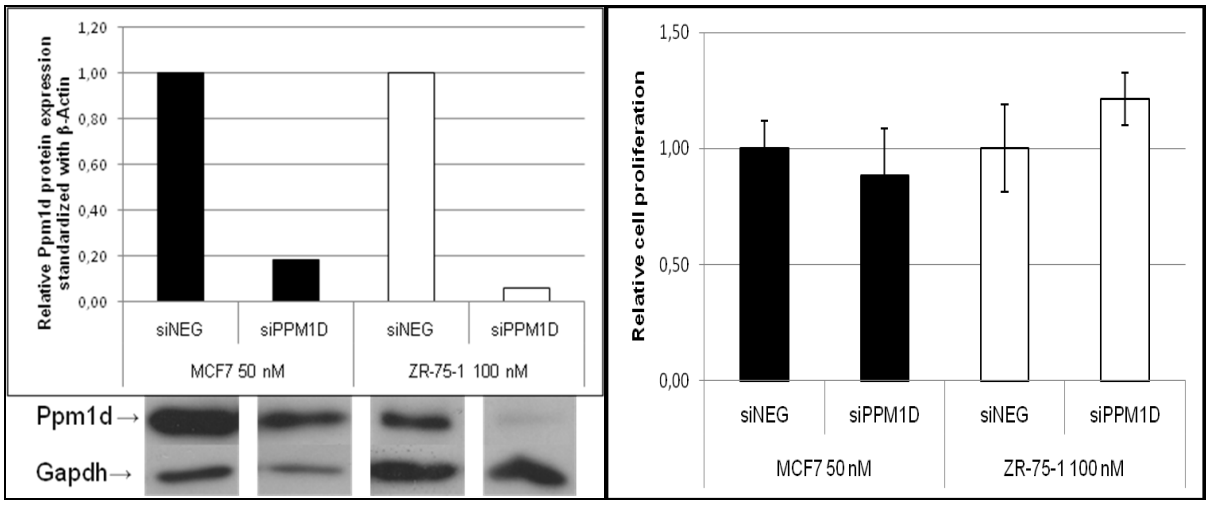
4.2. Setup and confirmation of proliferation assay: Knock down of PPM1D and PLK1 genes

In an effort to identify a putative oncogene on chromosomal region 8p12-p11 we used siRNA to downregulate the candidate genes expression. For these experiments breast cancer cell lines that had DNA amplification on the chromosomal region 8p12-p11 were used. To evaluate the genes ability to promote cancer growth we estimated cell proliferation after knocking down of the expression of the relevant gene. Proliferation was measured by counting cells or by a MTT-assay. These methods had not been used in our lab before and therefore we began by establishing the methods. As a positive control we used genes that are known to decrease cell proliferation when downregulated.

4.2.1. PPM1D was not sufficient to confirm methods reliably

Protein phosphatase Mg^{2+}/Mn^{2+} dependent, 1D (PPM1D) is a gene located on chromosome 17q23 a chromosomal region that has been found amplified in breast cancer (103). PPM1D codes for serine/threonine protein phosphatase, also known as Wip1. This protein has been shown to affect breast cancer cell lines when knocked down using RNA interference (104-106). MCF7 and ZR-75-1, overexpress the gene which seems to be the result of an amplification (104). Pärssinen and colleagues did a knock down on PPM1D using siRNA and found that the protein had significant effects on cell proliferation, in MCF7 and ZR-75-1 breast cancer cell lines (104).

We decided to use PPM1D as a positive control because this gene followed our model where the gene was amplified and both the mRNA and protein levels are increased (103, 104, 107). siRNA mediated knock down on the PPM1D gene was performed in MCF7 and ZR-75-1, in order to repeat Pärssinen's experiments (104).



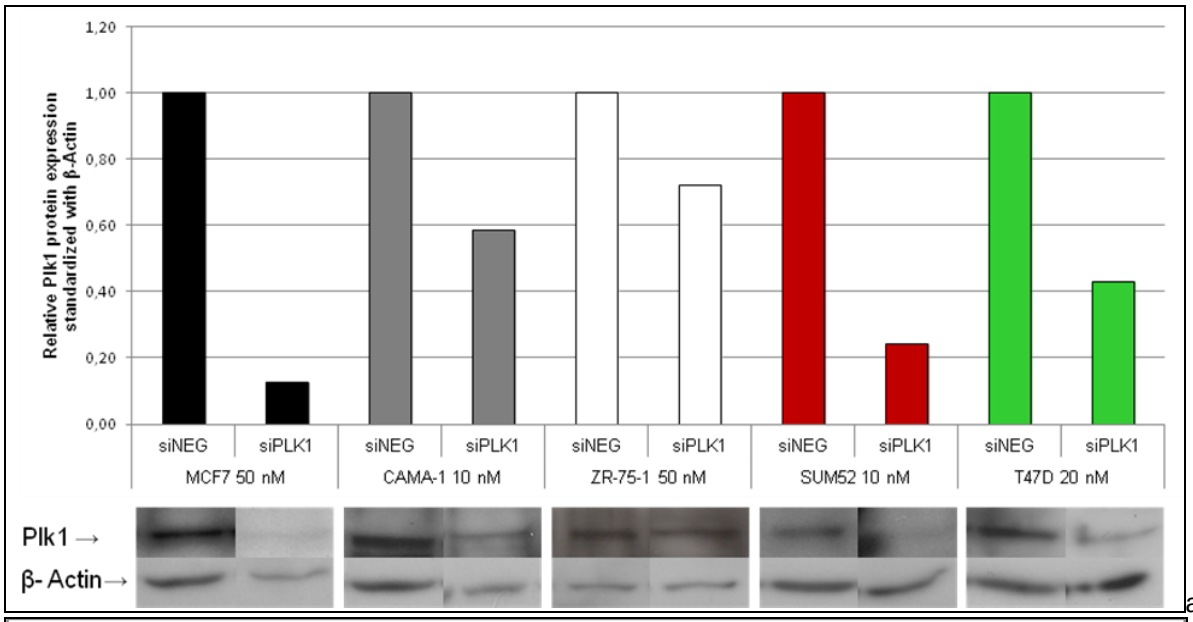
a) b)

Figure 8. PPM1D down regulation does not affect cell proliferation in MCF7 and ZR-75-1. Results from representative experiment. MCF7 (black columns) and ZR-75-1 (white columns) were treated with siRNA targeting no gene (siNEG) or PPM1D (siPPM1D) as indicated. MCF7 were harvested 72 hours after treatment and ZR-75-1 after 96 hours for protein and cell proliferation analysis. a) Protein lysates (5 µg/lane) were loaded on SDS-PAGE for electrophoresis and then transferred to PVDF membrane that was blotted with antibodies against Ppm1d (0.07µg/ml) and Gapdh (0.057 µg/ml). Ppm1d levels were standardized with Gapdh and cells treated with siNEG were set as 1. b) Cell proliferation of MCF7 was estimated with MTT-assay for MCF7 and by counting for ZR-75-1 after downregulation of PPM1D gene expression. Proliferation of cells treated with siNEG was set as 1, bars represent standard deviation. Experiments were done in triplicate.

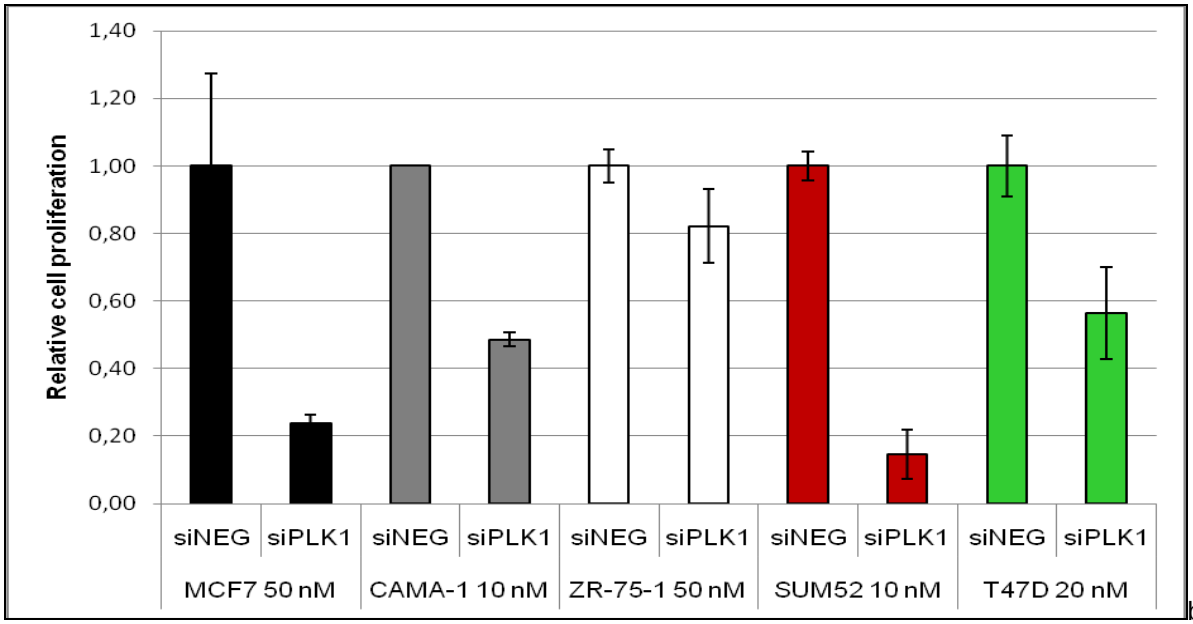
When MCF7 and ZR-75-1 were treated with siRNA targeting PPM1D, protein levels decreased to 18% and 6% respectively, as shown in Figure 8a. After knocking down the protein expression of Ppm1d, cell proliferation was estimated with MTT-assay, for MCF7, and cell counting, for ZR-75-1. Figure 8b shows that PPM1D down regulation did not result in decreased cell proliferation in either MCF7 or ZR-75-1. Even though the gene had been successfully downregulated and decreased cell proliferation had been observed by another group, we did not observe any effects (104).

4.2.2. PLK1 knock down resulted in convincing decrease in cell proliferation

Since the knockdown of PPM1D did not confirm the assay setup we decided to use another gene and we chose PLK1, (polo like kinase 1). PLK1 is important for all check points during the cell cycle (108). Cells in which PLK1 is downregulated only survive for few cell cycles before they undergo apoptosis (109, 110). PLK1 does not follow the same model as PPM1D and our target genes: that is DNA amplification followed by increased mRNA and protein expression. According to our previous experiments the knock down and its measurements were reliable, but we did not detect a decrease in cell proliferation (Figure 8). We used PLK1 downregulation to test if we were able to measure changes in cell proliferation after knock down. PLK1 gene expression was downregulated using siRNA in 5 different breast cancer cell lines, MCF7, CAMA-1, ZR-75-1, SUM52 and T47D. In all cases the knock down was a success, (Figure 9a), and in all cases the knock down resulted in a decreased proliferation (Figure 9b). The difference in changes in cell proliferation might be explained by knock down efficiency (Figure 9).



a)



b)

Figure 9. Knocking down PLK1 gene expression results in decreased cell proliferation. Results from representative experiment. MCF7 (black columns), CAMA-1 (grey columns), ZR-75-1 (white columns), SUM52 (red columns) and T47D (green columns), were treated with the indicated amount of siRNA, targeting either no gene (siNEG) or PLK1 (siPLK1). 72 hours (MCF7), 96 hours (CAMA-1, SUM52 and T47D) or 120 hours (ZR-75-1) after treatment cells were harvested for protein expression or cell proliferation estimation. a) Cells protein lysates (5 µg/lane) were loaded on SDS-PAGE for electrophoresis and then transferred on PVDF membrane that was blotted with antibodies against Plk1 (0.03 µg/ml) and β-actin (0.5 µg/ml). Protein expression was quantified as described in materials and methods. β-Actin expression was used to standardized Plk1. Protein expression in cells treated with siNEG was set as 1. The figure below shows an example of western blots. b) Cell proliferation was measured in MCF7, CAMA-1 and ZR-75-1 with MTT-assay but in SUM52 and T47D proliferation was measured by cell counting. The proliferation of cells treated with siNEG was set as 1. Standard deviations are shown with bars. Experiments were done in triplicate.

These results show that our methods can detect changes in cell proliferation after treatment. Knock down of PLK1 was used in every experiment for verification from now on.

4.3. The search for putative oncogenes within amplicon 8p12-p11

For a gene to reach our list of candidate target genes (Table 4) it had to fulfill a few criteria we set: First, the gene had to be located within the minimal amplified region, that is the region that is always amplified when amplification on 8p12-p11 has occurred; second, the mRNA expression of the gene had to be elevated when the gene was amplified and finally, the final product of the gene, the protein levels, had to be elevated as well (described in Introduction).

Table 4. The candidate target genes on the amplified region at 8p12-p11. A list of genes located within the minimal amplicon at 8p12-p11 which fulfilled specific criteria to be considered possible target genes of the amplification. Genes that were tested for effects on cell proliferation in this thesis are listed in bold. Y = yes, N = no, NA = not available (16).

Genes – listed in chromosomal order	Size (kDa)	DNA – mRNA correlation	mRNA – protein correlation	DNA – protein correlation
ERLIN2	**38**	**Y**	**Y**	**Y**
PROSC	**30**	**Y**	**Y**	**Y**
BRF2	50	Y	NA	Y
RAB11FIP1	83	Y	Y	Y
	42		NA	N
ASH2L	**82**	**Y**	**N***	**Y**
LSM1	**17**	**Y**	**Y**	**Y**
DDHD2	95		NA	Y
	72		NA	N
	36	Y	N	N

*The aGEx results were not of sufficient quality for correlation calculations.

To approach the question whether the candidate genes were essential for cancer cell growth we downregulated gene expression in selected breast cancer cell lines and measured cell proliferation. The cell lines we chose for this purpose are listed in Table 5. We chose cell lines based on the status of our target genes on 8p12-p11. DNA status of the genes was based on published data, as indicated in Table 5, but protein expression is based on our own measurements, done with western blot. Like breast tumors have been classified into five different subtypes, many breast cancer cell lines have undergone the same process. Classification of breast tumors is mainly based on global mRNA expression but as said before these results have also been verified by immunohistochemistry (20, 32). The same approach has been used to classify breast cancer cell lines (34). Cancer cell lines are known to carry genomic alterations that are not found in tumors which could interfere with the classification (111). In addition some subtypes can be hard to distinguish between such as luminal A and luminal B but these subtypes are often classified based on the expression levels of ER, PgR and related genes (34, 112). For breast cancer cell lines the luminal subtypes, A and B are often pooled into one subtype; the luminal subtype.

Table 5. Cell lines used for knock down and cell proliferation studies. A list of the cell lines used to knock down a putative oncogene and measure the effect on cell proliferation. The status of the genes we tested in this thesis according to published data and protein levels in each cell line according to western blots performed in our lab as well the most resembling breast cancer subtype. Amp = DNA amplification (ratio > 1.8), Gain = Gained/increased DNA levels (ratio 1.2-1.8), Norm = normal DNA levels (ratio 0.8 – 1.2) and loss = decreased DNA levels (ratio < 0.8). Protein expression is estimated compared to D492-3D, were the protein expression is set as 1. Lum = luminal subtype, Lum A = luminal A subtype.

		CAMA-1	MCF7	SUM52	T47D*	ZR-75-1
ERLIN2	DNA	Amp	Loss	Amp	Gain	Amp
	Protein	19.4	0.7	1.8	12.7	12.9
PROSC	DNA	Amp	Loss	Amp	Gain	Amp
	Protein	6.1	2.3	7.6	4.7	4.6
ASH2L	DNA	Amp	Loss	Amp	Norm/Gain	Amp
	Protein	10.5	3.1	3.7	4.1	17.6
LSM1	DNA	Amp	Loss	Amp	Norm/Gain	Amp
	Protein	2.4	1.0	2.0	1.8	3.3
Subtype		Lum	Lum A	Lum	Lum	Lum A
References		(96, 97)	(34, 95, 98)	(27, 41, 94, 96, 97)	(94, 95, 97)	(34, 95-98)

*One article describes translocation between chromosome 8 and chromosome 14, observed with FISH, where both ASH2L and LSM1 are lost in the process. Other published data do not describe this, possible explanation is that aCGH is used to measure DNA content (95, 97).

4.3.1. Erlin2 knock down in breast cancer cell lines does not affect cell proliferation

The first gene we tested was ERLIN2, that codes for the protein ER lipid raft associated 2, a protein that plays a role in degradation of IP_3 receptors after their activation (59, 62, 64). Our hypothesis was that with higher levels of Erlin2 more IP_3 receptors were sent to degradation. This resulted in lack of IP_3 receptors in the cells and then cells avoidance of apoptosis.

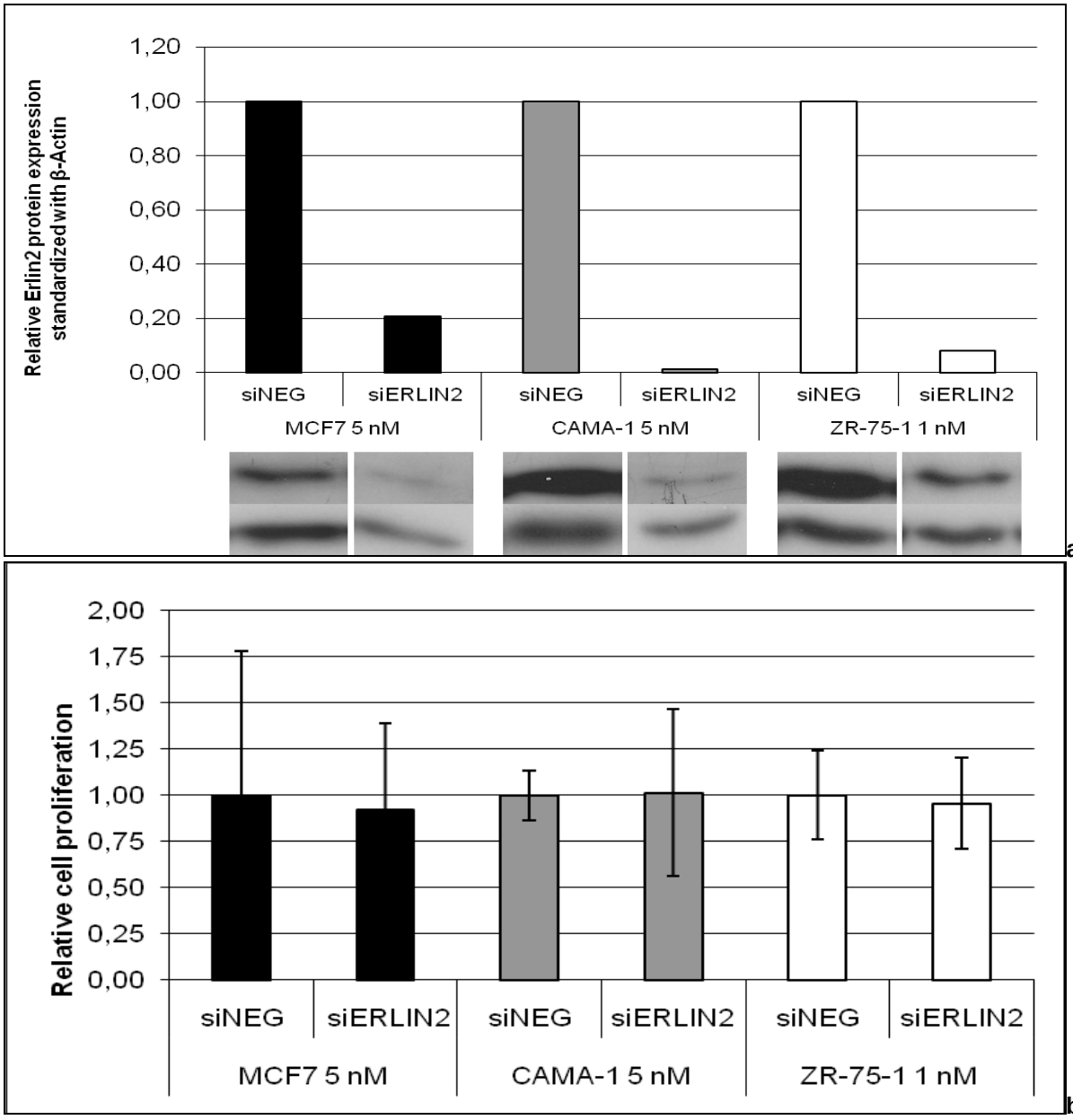


Figure 10. Erlin2 knock down does not affect cell proliferation. Results from representative experiment. MCF7 (black columns), CAMA-1 (grey columns) and ZR-75-1 (white columns) were treated with the indicated amount of siRNA targeting either no gene (siNEG) or ERLIN2 (siERLIN2). 72 hours after treatment the cells were harvested for protein expression or cell proliferation estimation. a) Cells protein lysates (5 µg/lane) were loaded on SDS-PAGE for electrophoresis and then transferred to PVDF membrane that was blotted with antibodies against Erlin2 (0.5 µg/ml) and β-Actin (0.5 µg/ml). Protein expression was quantified as described in materials and methods and standardized with β-Actin. Protein expression in cells treated with siNEG was set as 1. Western blot results are shown below. b) Cell proliferation was measured with MTT-assay. Proliferation is shown relative to cells treated with siNEG which was set as 1. Standard deviation are shown with bars. Experiments were done in triplicate.

To test if ERLIN2 was the 8p12-p11 amplicons target gene and an oncogene, we downregulated its expression in three breast cancer cell lines, MCF7, ZR-75-1 and CAMA-1. MCF7 harbors loss of DNA on chromosomal region 8p12-p11 but both ZR-75-1 and CAMA-1 harbor amplification (Table 5). In addition CAMA-1 and ZR-75-1 overexpress the Erlin2 protein. Erlin2 protein expression levels, in CAMA-1 and ZR-75-1, are 19.4 and 12.9 fold increased, respectively, when compared to normal cell D492-3D (Figure 11). We aimed to downregulate Erlin2 expression to the same or below the expression levels observed in the D492-3D cells.

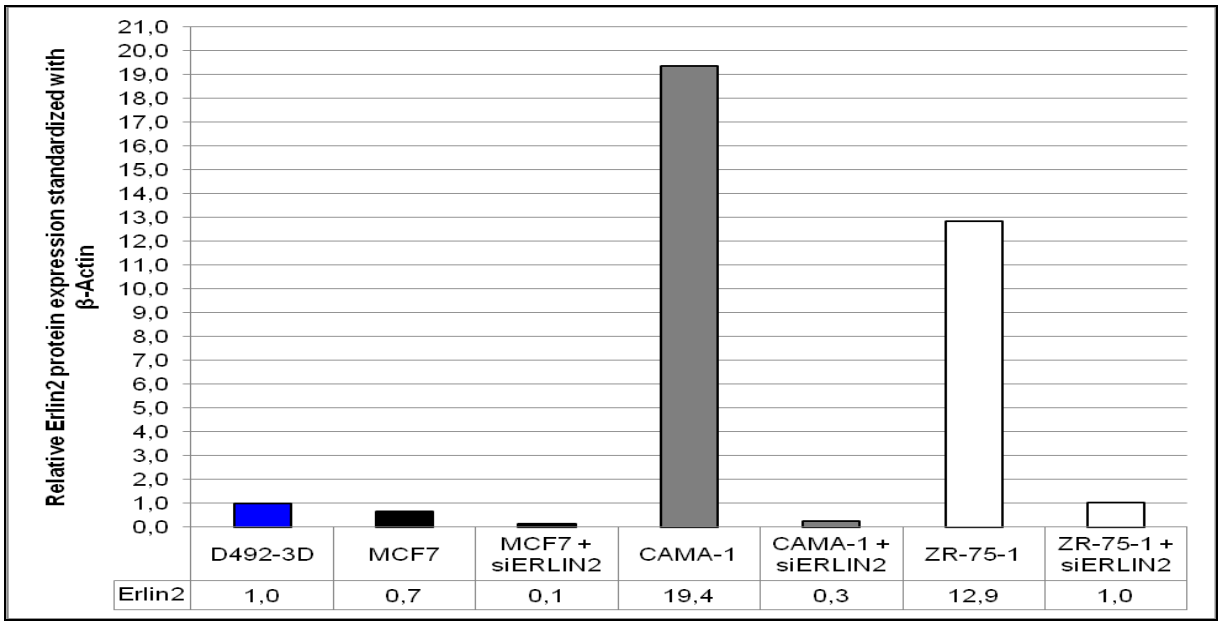


Figure 11. Erlin2 protein expression in normal breast cell line (D492-3D) and three breast cancer cell lines. Protein lysates (5 µg/lane) from D492-3D (blue), MCF7 (black), CAMA-1 (grey) and ZR-75-1 (white) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against Erlin2 (0.5 µg/ml) and β-Actin (0.5 µg/ml). Protein levels were quantified as described in materials and methods and standardized with β-Actin. siRNA targeting either no gene (siNEG) or ERLIN2 (siERLIN2) were transfected into MCF7 (5 nM), CAMA-1 (5 nM) and ZR-75-1 (1 nM). . 72 hours after treatment cells were harvested and protein expression analysis was done as described above. The table below shows Erlin2 expression, relative to Erlin2 levels in D492-3D, which was set as 1.

Knock down of Erlin2 was successful (Figure 10a), Erlin2 protein levels in CAMA-1 and ZR-75-1 reached the same or below the protein expression as D492-3D, respectively (Figure 11). The effects of ERLIN2 knock down on cell proliferation were measured with MTT-assay, and no effects were detected (Figure 10b). Based on these results we were able to exclude ERLIN2 as a gene affecting cell proliferation in the cell lines tested here.

4.3.2. LSM1 knock down in breast cancer cell lines does not affect cell proliferation

LSM1 plays a role in mRNA processing in cytoplasm and P-body formation (113, 114), (115). Knock down of LSm1 was done in three breast cancer cell lines, MCF7, CAMA-1 and ZR-75-1 (Figure 12). CAMA-1 and ZR-75-1 harbor amplification of this particular gene, but MCF7 has lost a copy of this gene (Table 5). We estimated LSm1 protein expression in all three cell lines relative to expression in the normal cell line D492-3D (Figure 13). Downregulation of the gene in MCF7, CAMA-1 and ZR-75-1 decreased LSm1 levels down to 6%, 4% and 16% from the original protein levels in each cell line, respectively (Figure 12a).

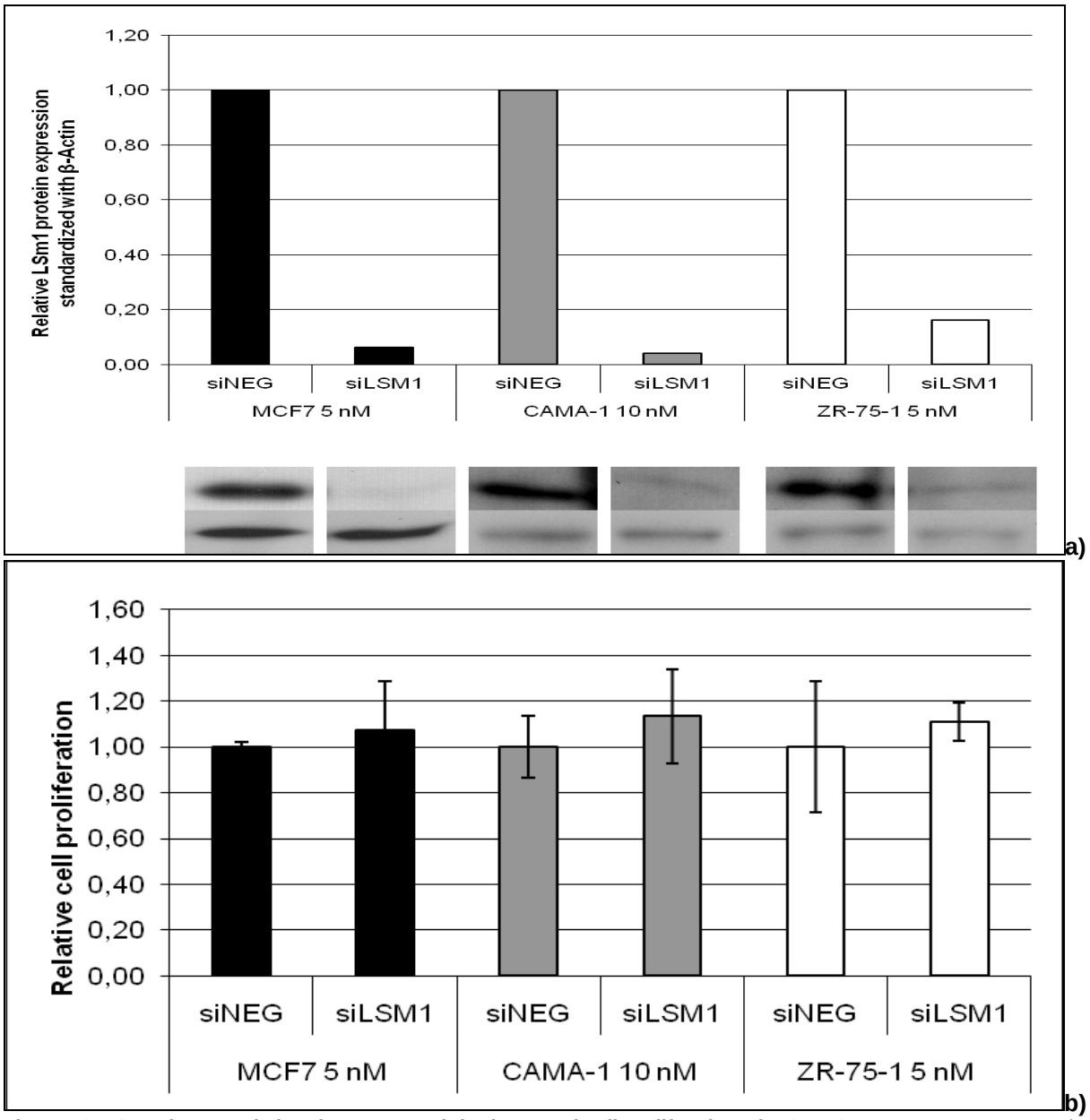


Figure 12. LSm1 downregulation does not result in decreased cell proliferation of MCF7, CAMA-1 or ZR-75-1. Results from representative experiment. Three cell lines MCF7 (black columns), CAMA-1 (grey columns) and ZR-75-1 (white columns) were treated with siRNA targeting either no gene (siNEG) or LSM1 (siLSM1), siRNA concentration is indicated below each cell line name. 72 hours after treatment cells were harvested for protein expression or cell proliferation estimation. a) Protein lysates (5 µg/lane) were loaded on SDS-PAGE and then transferred to PVDF membrane that was blotted with antibodies against LSm1 (1.0 µg/ml) and β-Actin (0.05 µg/ml). Protein expression was quantified as described in materials and methods and standardized with β-Actin. Protein levels in cells treated with siNEG was set as 1. Western blot results shown below the graph. b) Cell proliferation was measured using MTT-assay. Proliferation is shown relative to cells treated with siNEG, which was set as 1. Standard deviation is shown with bars. Experiments were done in triplicate.

We sought to decrease LSm1 levels in CAMA-1 and ZR-75-1 down to or below the protein levels in D492-3D. As shown in Figure 13, MCF7 and D492-3D express the same amount of LSm1 but both CAMA-1 and ZR-75-1 express higher levels of LSm1, or 2.43 fold and 3.33 fold levels, respectively. When CAMA-1 and ZR-75-1 were treated with siRNA targeting LSM1 the protein levels decreased below the LSm1 levels in D492-3D (Figure 13).

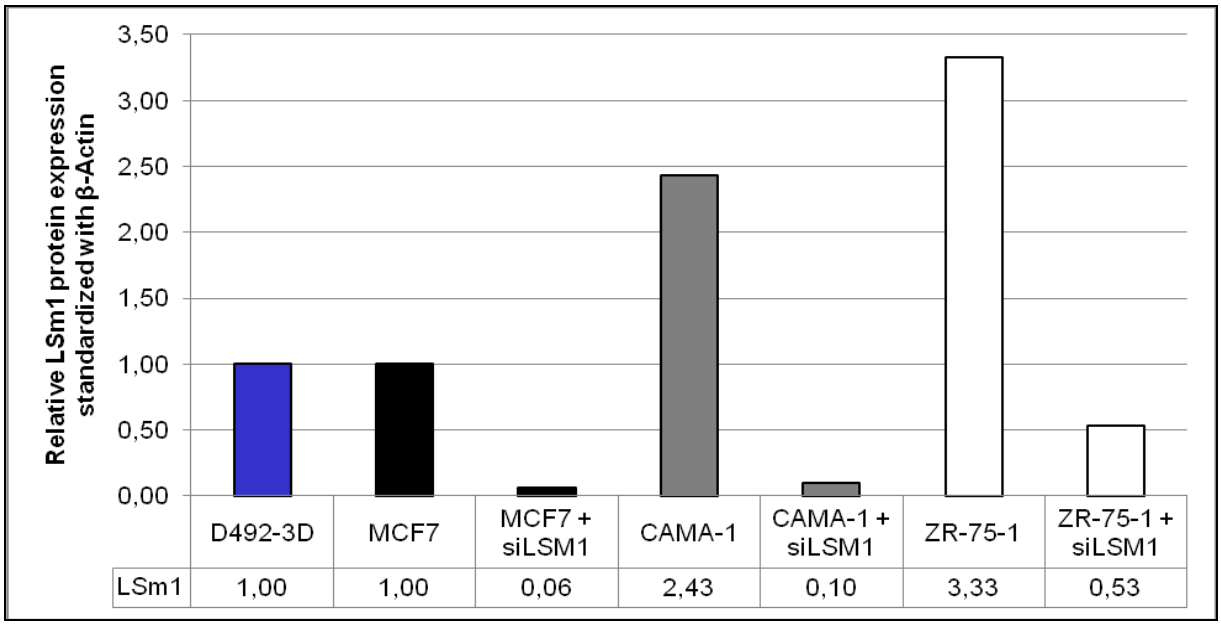
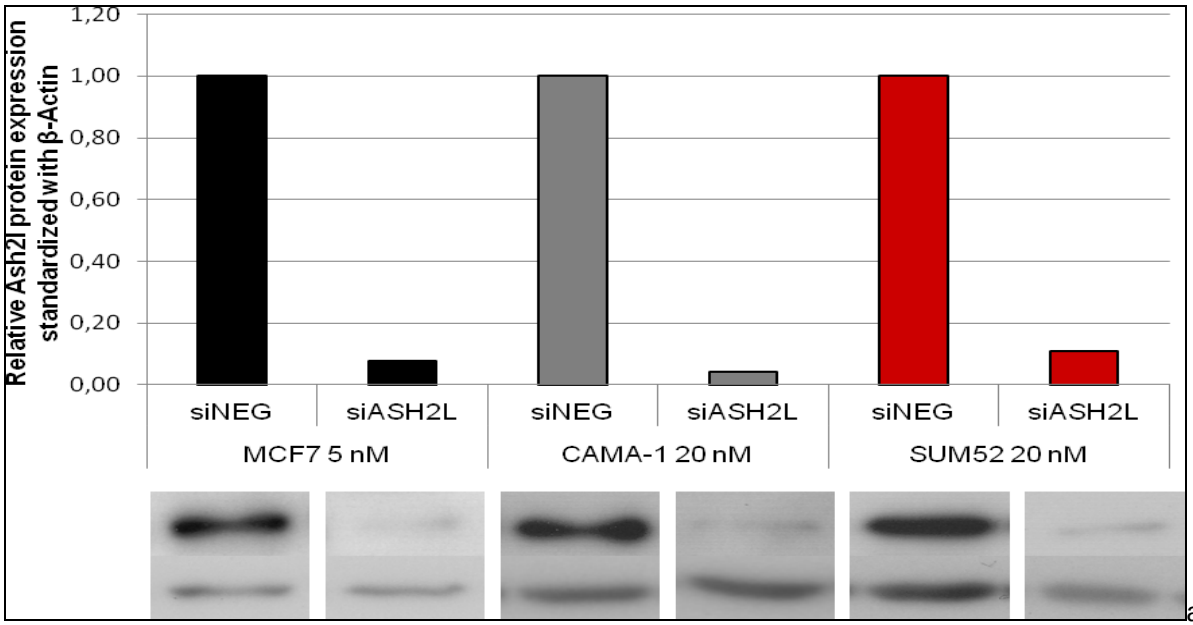


Figure 13. LSm1 protein expression in normal cell line, D492-3D, and breast cancer cell lines, MCF7, CAMA-1 and ZR-75-1. Protein lysates (5 µg/lane) from D492-3D (blue), MCF7 (black), CAMA-1 (grey) and ZR-75-1 (white) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against LSm1 (1.0 µg/ml) and β-Actin (0.5 µg/ml). Protein levels were quantified as described in materials and methods and standardized with β-Actin. MCF7, CAMA-1 and ZR-75-1 were treated with siRNA targeting either no gene (siNEG) or LSM1 (siLSM1) 5 nM, 10 nM and 5 nM, respectively. 72 hours after treatment cells were harvested and protein expression analysis was done as described above. The table below shows LSm1 expression, relative to LSm1 levels in D492-3D, which was set as 1.

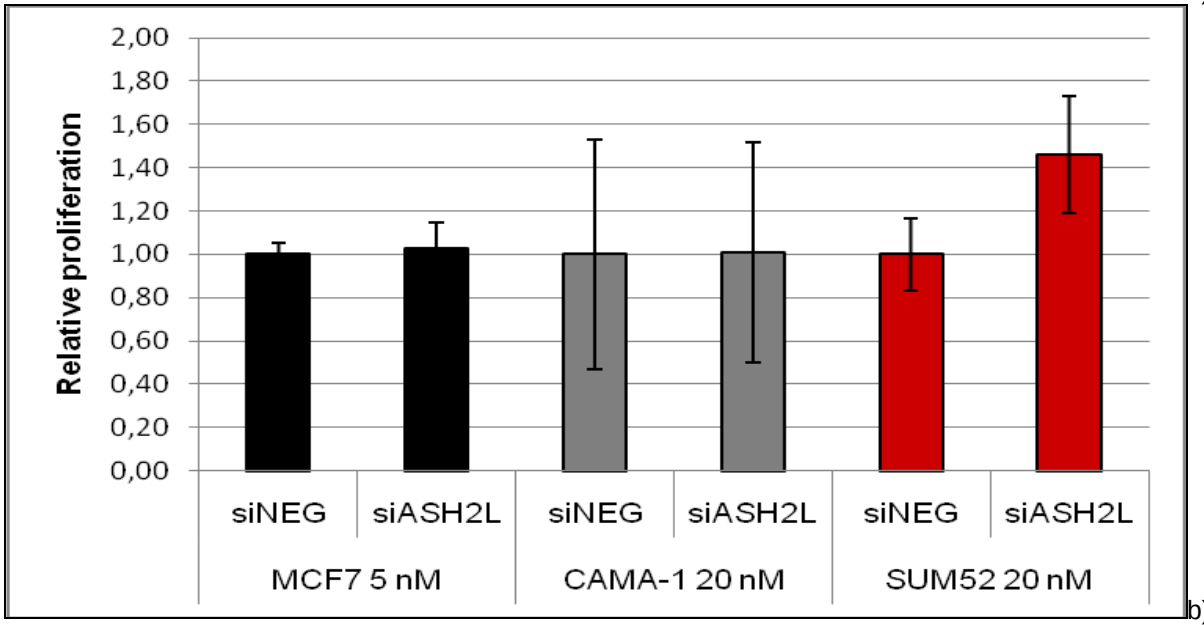
Downregulating LSM1 did not affect cell proliferation in CAMA-1 or ZR-75-1 (Figure 12b). These results do not support LSM1 as the amplicon's target gene.

4.3.3 ASH2L knock down does not affect cell proliferation in breast cancer cell lines

ASH2L codes for methyltransferase protein that facilitates increased gene expression (82, 83). We hypothesized that Ash2l was by altering gene expression, affecting cell proliferation when amplified and overexpressed in breast tumors. To test this we downregulated the ASH2L gene in three breast cancer cell lines: MCF7, CAMA-1 and SUM52.



a)



b)

Figure 14. Downregulation of Ash2l does not affect cell proliferation in MCF7, CAMA-1 or SUM52. Results from representative experiment. Three breast cancer cell lines, MCF7 (black columns), CAMA-1 (grey columns) and SUM52 (red columns), were treated with the indicated amount of siRNA targeting no gene (siNEG) or ASH2L (siASH2L). 96 hours after treatment cells were harvested for protein expression and cell proliferation analysis. a) Protein lysates (10 µg/lane) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against Ash2l (2.0 µg/mg) and β-Actin (0.5 µg/ml). Protein expression was quantified as described in materials and methods and standardized with β-Actin. Ash2l expression in cells treated with siNEG was set as 1. Western blots shown below. b) Cell proliferation was tested by cell counting. Cell proliferation of cells treated with siNEG was set as 1. Standard deviations are shown with bars. Experiments were done in triplicate.

ASH2L knock down was successful in all three cell lines, MCF7, CAMA-1 and SUM52, protein decrease reached 8%, 4% and 11% of the levels in each cell line, respectively, (Figure 14a). Knocking down Ash2l protein expression did not result in decreased cell proliferation in either CAMA-1 or SUM52 (Figure 14b). Both of those cell lines harbor amplification of this particular gene (Table 5) and express higher levels of Ash2l than normal cell D492-3D (Figure 15). We estimated that MCF7 should not gain any benefits from ASH2L, but MCF7 carries a loss on the 8p12-p11 region (Table 5). When we measured the Ash2l protein expression, we found that despite of potential loss, of chromosomal

region 8p in MCF7, the cell line was still expressing more protein than D492-3D (Figure 15 and Table 5).

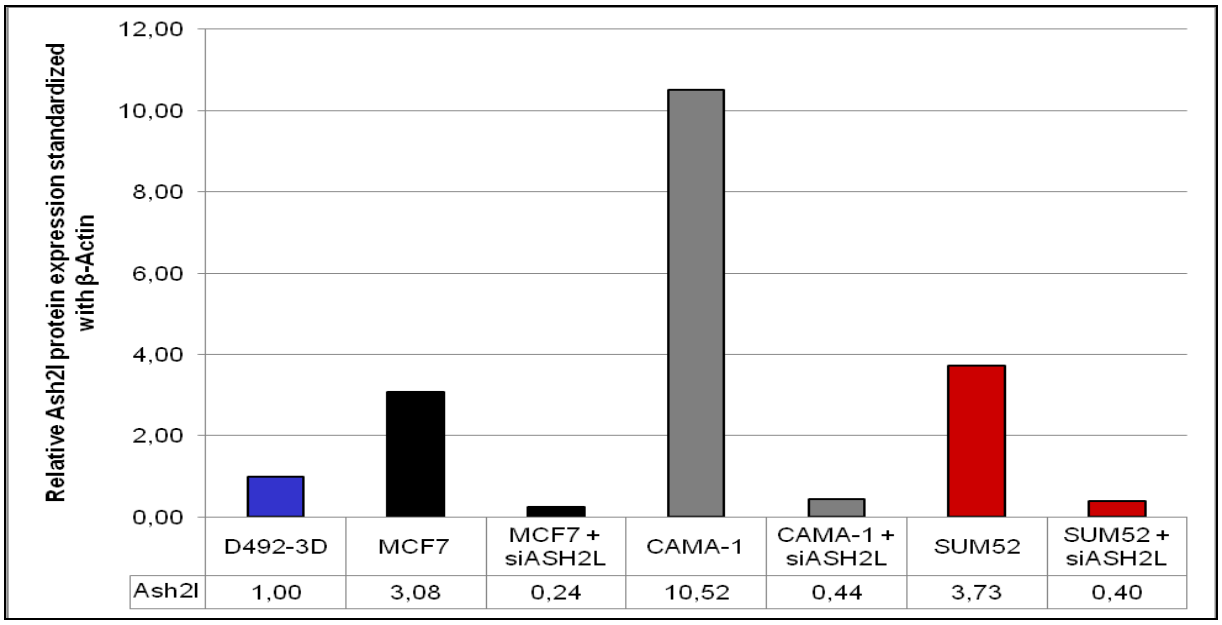


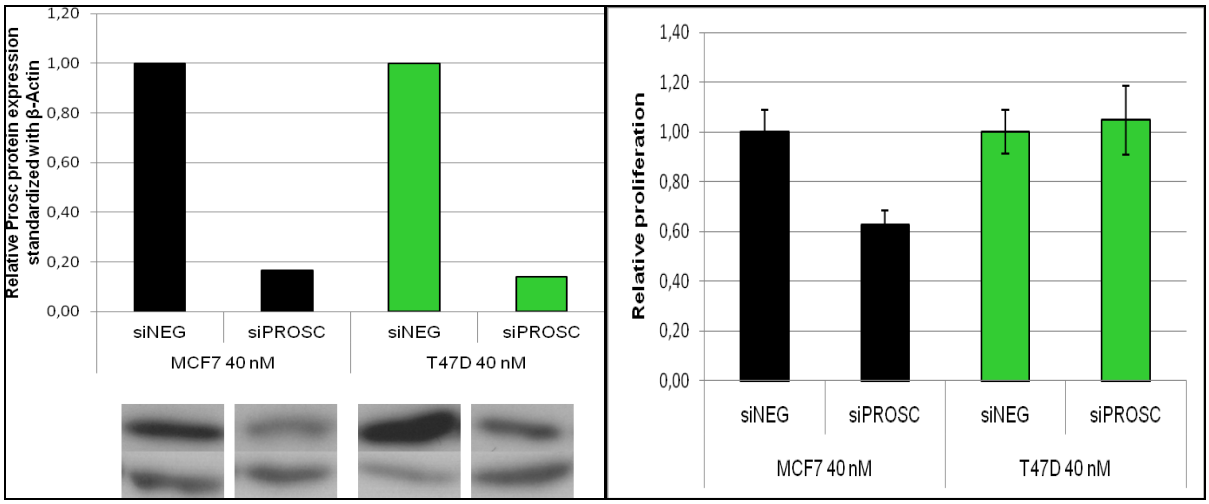
Figure 15. ASH2L protein expression in one normal breast cell line, D492-3D, and three breast cancer cell lines, MCF7, CAMA-1 and SUM52. Protein lysates (10 µg/lane) from D492-3D (blue), MCF7 (black), CAMA-1 (grey) and SUM52 (red) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against Ash2l (2.0 µg/ml) and β-Actin (0.5 µg/ml). Protein levels were quantified as described in materials and methods and standardized with β-Actin. MCF7, CAMA-1 and SUM52 were treated with 5 nM (MCF7) or 20 nM (CAMA-1 and SUM52) siRNA targeting no gene (siNEG) or ASH2L (siASH2). 96 hours after treatment cells were harvested and protein expression was analysed as described above. The table below shows Ash2l protein levels relative to Ash2l protein levels in D492-3D, which was set as 1.

In both cell lines, CAMA-1 and SUM52, we were able to decrease Ash2l levels below the D492-3D levels (Figure 15). Based on these results we concluded that Ash2l does not contribute to cell proliferation in the breast cancer cell lines CAMA-1 and SUM52. These results do not support the hypothesis that ASH2L is the target gene of the amplification on 8p in breast tumors.

4.3.4 PROSC knock down does not affect cell proliferation in breast cancer cell lines

The next gene tested was PROSC (Table 4). The complex pattern of the 8p12-p11 amplicon is represented in the breast cancer cell line T47D which harbors gain on the amplicon (95, 97). Mackay and colleagues describe gains over the 8p12-p11 region, which is measured using aCGH. The BACs used in their experiments detect seven genes from the minimal amplicon defined in our lab (16). These genes are, ZNF703, GRP124, EIF4EBP1, ASH2L, PPAPDC1B, WHSC1L1 and FGFR1 (95). In addition Pole and her colleagues used FISH to measure the DNA abundance on this region. They found that the cell line also carries a translocation from the p-arm of chromosome 8 to chromosome 14 (97). In the process of this translocation a part of the 8p12 region that contains ASH2L, STAR, LSM1 and BAG4 is lost. The genes on either side are however found in elevated copies (97). Based on these data we decided to test PROSC knock down in T47D, based on the fact that the cell line

harbored gain on PROSC in higher levels than both ASH2L and LSM1 (Table 5) (97). For comparison PROSC was also downregulated in a cell line carrying loss on the 8p region, MCF7 (Table 5).



a) b)

Figure 16. Prosc knock down affects cell proliferation in MCF7 but not in T47D breast cancer cell lines. Two breast cancer cell lines MCF7 (black columns) and T47D (green columns) were treated with the indicated amount of siRNA targeting either no gene (siNEG) or PROSC (siPROSC). 96 hours after treatment cells were harvested for protein expression and cell proliferation estimation a) Protein lysates (5 µg/lane) were loaded on SDS-PAGE and then transferred to PVDF membrane that was blotted with antibodies against PROSC (0.33 µg/ml) and β-Actin (0.5 µg/ml). Protein expression was quantified as described in materials and methods and standardized with β-Actin. Prosc expression in cells treated with siNEG was set as 1. Western blots are shown below. b) Cell proliferation was measured by cell counting. Proliferation of cells treated with siNEG was set as 1. Standard deviations are shown with bars.

PROSC knock down did not affect the breast cancer cell line T47D (Figure 16). Unexpectedly the knock down reduced cell proliferation in MCF7. An effect on MCF7 was not expected since the cell line harbors a loss of the 8p12-p11 amplicon. These experiments have not been repeated yet so the results should not be interpreted until they have been further verified.

We also estimated Prosc expression in these breast cancer cell lines compared to normal breast line, D492-3D, (Figure 17). Downregulation of PROSC protein expression in both MCF7 and T47D reached below the protein levels in D492-3D. MCF7 expressed more Prosc than D492-3D. This was surprising considering that MCF7 carries a loss on chromosomal region 8p12-p11 (Table 5). Like for the Ash2l protein expression, shown in Figure 15, these results might indicate that Prosc protein expression is increased in breast cancer cell lines independent from DNA amplification.

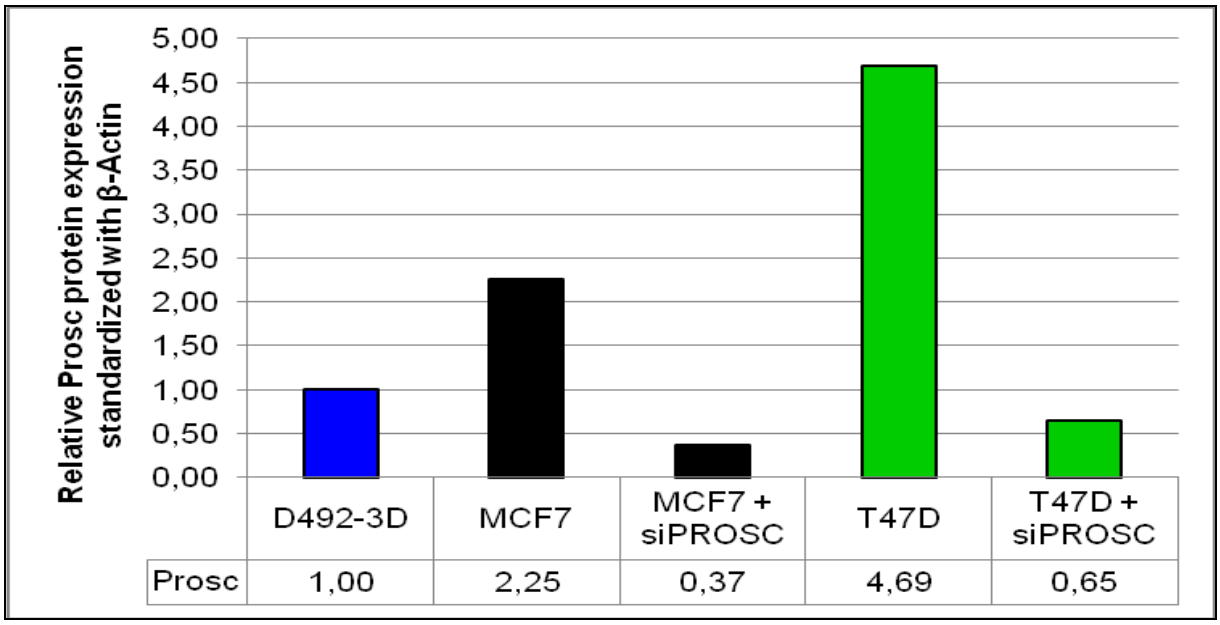


Figure 17. Prosc protein expression in breast cancer cell lines MCF7 and T47D exceeds Prosc levels in D492-3D normal breast cell line. Protein lysates (5 µg/lane) from D492-3D (blue), MCF7 (black) and T47D (green) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against Prosc (0.33 µg/ml) and β-Actin (0.5 µg/ml). Protein levels were quantified as described in materials and methods and standardized with β-Actin. MCF7 and T47D were treated with 40 nM siRNA targeting no gene (siNEG) or ASH2L (siASH2). 96 hours after treatment cells were harvested and protein expression was analysed as described above. The table below shows Prosc protein levels relative to Prosc protein levels in D492-3D, which was set as 1.

Even though Prosc knock down seemed to affect cell proliferation in MCF7 breast cancer cell line this experiment has only been done once and has to be repeated before any conclusions can be made. Knock down in T47D had no effect on cell proliferation, which is not in support of PROSC being the target gene of the 8p12-p11 amplicon in breast cancer.

4.3.5. ASH2L, LSM1 and PROSC knocked down simultaneously

ASH2L, LSM1 and PROSC showed good correlation between DNA status and protein expression. In addition significant coexpression of ASH2L, LSM1 and PROSC was observed in the breast tumor samples, according to western blot analyses (16). Based on these results, we postulated that the three genes might be in cooperation with each other, increasing carcinogenesis. Other groups have also implied that genes located 8p12-p11 might be working together in carcinogenesis (27, 41, 45). We therefore sought to test if downregulation of these three genes could decrease cell proliferation in breast cancer cell lines. To address this question we did triple knock down. ASH2L, LSM1 and PROSC were downregulated in four cell lines, MCF7, CAMA-1, SUM52 and T47D and cell proliferation was measured.

The knock down was only successful in one cell line, CAMA-1. CAMA-1 harbors amplification of all three genes (Table 5).

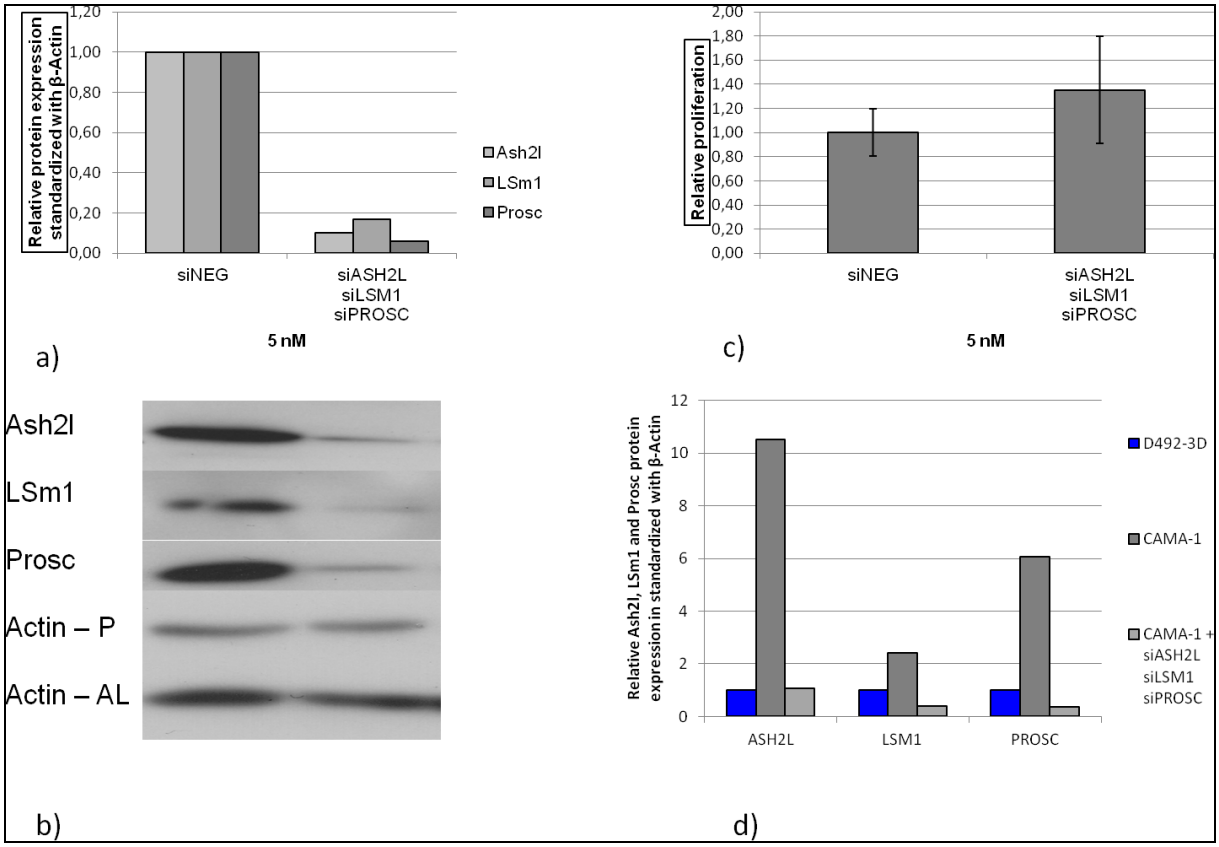


Figure 18. Downregulating three genes at once, ASH2L, LSM1 and PROSC, does not alter cell proliferation in CAMA-1. Results from representative experiment. CAMA-1 was treated with the indicated amount of siRNA targeting no gene (siNEG) or ASH2L (siASH2L), LSM1 (siLSM1) and PROSC (siPROSC), simultaneously. 96 hours after treatment the cells were harvested for protein expression and cell proliferation analysis. a) A graph depicting relative protein abundance from the gels shown in b). b) Protein lysates (5 µg/lane (Prosc) and 10 µg/lane (Ash2l and LSm1)) were loaded on SDS-PAGE and then transferred to PVDF membrane which was blotted with antibodies against Ash2l (2.0 µg/ml), LSm1 (1.0 µg/ml), Prosc (0.33 µg/ml) and β-Actin (0.5 µg/ml). Protein expression was quantified and standardized with β-Actin. LSm1 and Ash2l were on the same blot and thus the same Actin used for standardization (Actin – AL) but Prosc was on a separate gel (Actin – P). Protein expression in cells treated with siNEG was set as one. c) Cell proliferation was estimated with cell counting. Cell proliferation of cells treated with siNEG was set as one. Standard deviations are shown with bars. Experiments were done in triplicate. d) Protein lysates (5 µg/lane (Lsm1 and Prosc) and 10 µg/lane (Ash2l)) from normal cell line D492-3D (blue columns), CAMA-1 treated with siNEG (dark grey columns) and CAMA-1 treated with siASH2L, siLSM1 and siPROSC (light grey columns), were loaded on SDS-PAGE, transferred to PVDF membrane that was blotted with antibodies against Ash1l (2.0 µg/ml), LSm1 (1.0 µg/ml), Prosc (0.33 µg/ml) and β-Actin (0.5 µg/ml). Protein expression was quantified and standardized with β-Actin. Ash2l, LSm1 and Prosc protein expression in D492-3D was set as one.

The aim was to reduce protein expression of all the three genes in CAMA-1 down to the same levels or below the levels observed in D492-3D. Figure 18a and b shows the results from an experiment were all three genes, ASH2L, LSM1 and PROSC were successfully downregulated in CAMA-1. Downregulation in CAMA-1 reached the same protein expression as in D492-3D for all three genes, and both LSm1 and Prosc levels reach well below D492-3D levels (Figure 18d). Even though the knock down was good the effects on cell proliferation were none (Figure 18c).

Based on these results we concluded that ASH2L, LSM1 and PROSC are not major impact genes in cell proliferation in CAMA-1. That does not exclude these genes as possible oncogenes in other cell lines that might harbor different genetic background, such as SUM52, ZR-75-1 and T47D.

5. Discussions

In this thesis we have shown that increased copy numbers of the LETM2 and PPAPDC1B genes do not lead to increased protein expression in breast tumors. Even though the breast cancer cell lines do display elevated protein levels, the protein abundance in breast tumors does not fit our criteria to be listed as target genes of the amplification at 8p12-p11 (Figure 1 and 6). Combining this study with a study previously done in our lab by Berglind Ósk Einarsdóttir, the list thus counts seven putative oncogenes (16): ERLIN2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1 and DDHD2. Using siRNA to downregulate expression of ERLIN2, LSM1, ASH2L and PROSC in breast cancer lines we have shown that these genes do not affect the cell proliferation in the cell lines tested (Figure 10, 12, 14 and 16). These results do not support the hypothesis that those genes might play a role in cancer growth. However, it does not exclude these genes as target genes of the 8p12-p11 amplicon since the effect of their upregulation could be related to something else than cancer growth.

5.1. Letm2 exclusion only based on two isoforms

The LETM2 gene was excluded as the amplicon target gene based on available data regarding the various isoforms the gene codes for. Letm2 did not show correlation between DNA amplification and increased protein expression. Based on the CCDS database the LETM2 gene should only code for one isoform of the protein that weighs 45 kDa (102). According to another database the LETM2 gene codes for nine different transcripts that could be translated into proteins (101). In addition the third database, uniprot (http://www.uniprot.org/, release 2011_11), has four isoforms listed (116). As shown in Figure 1 the antibody detects a signal at 45 kDa and 56 kDa molecular weight. The 56 kDa signal is stronger than the 45 kDa signal, and according to the antibody product information the antibody should recognize protein around 56 kDa. This gives evidence for the fact that the 56 kDa signal is indeed the Letm2. However the 45 kDa signal was also quantified during the process of this thesis for it is likely to be a different isoform of the Letm2 protein. To be able to exclude the Letm2 gene as a putative candidate gene, one would have to verify the existence of other isoforms of the protein and test if their expression is elevated.

Qualified mRNA data for the LETM2 gene were only available for few samples. Therefore correlation calculations between mRNA and DNA or protein were not done. The DNA data show however that the LETM2 is amplified in our tumor samples (16). The bacterial artificial chromosomes (BACs) that were used to measure DNA levels in aCGH most often detect more than one gene, so even if our measurements showed that LETM2 was amplified that might be due to other genes located on the same BACs. LETM2 was localized on a BAC that also contains WHSC1L1, FGFR1 and a gene called FLJ43582 or C8orf86, that codes for an unknown protein product. Wolf- Hirschhorn syndrome candidate 1 like 1 (WHSC1L1) along with a neighboring gene, PPAPDC1B, has been indicated as the driver gene of the 8p12-p11 amplicon (39) which was further supported in cell-based studies by Yang and colleagues (45). The WHSC1L1 gene was not included on our list of putative target genes (16). Fibroblast growth factor receptor 1 (FGFR1) was also located on this particular BAC. The Fgfr1 seems to be a good candidate protein for a tumor to overexpress, based on the proteins function. However

studies have shown that the FGFR1 gene is unlikely to be the amplicons target gene (16, 39). The FGFR1 overexpression has been linked to lobular breast cancer (117, 118). Overexpression of FGFR1 results in decreased metastatic disease free survival in patients and also increased colony formation in breast cell lines (110, 118). In the correlation study performed in our lab, where FGFR1 amplification and overexpression was measured the protein levels were not increased even though the gene was amplified (16). This might be due to lack of lobular tumors in our database, but the gene was found amplified like the LETM2 gene, located on the same BAC (16).

However, other studies support our results. LETM2 has not been identified as a candidate oncogene or the target gene of the amplicon even though the gene has been found amplified and with elevated mRNA levels in breast cancer (41, 43). The gene has also been eliminated based on lack of elevated mRNA expression even though the DNA copy number was increased (39). Our study is the first to show that Letm2 protein expression is not increased when the gene is amplified.

5.2. Ppapdc1B behaves like a tumor suppressor protein

PPAPDC1B is often amplified in breast cancer and simultaneously overexpressed at the mRNA level (39, 41, 43). In fact, the gene has been characterized as an oncogene and the driver gene for the 8p12-p11 amplification by Isabelle Bernard-Pierrot and her colleagues (39). In their study they used siRNA to downregulate the gene in breast cancer cell lines. The downregulation resulted in decreased cell proliferation, induced apoptosis along with decreased focus formation and growth in soft agar. When tumors, established in mice, were treated with siRNA targeting PPAPDC1B they decreased significantly compared to tumors treated with scrambled siRNA (39). Our results are contradictory to the results of Bernard-Pierrot et al. According to our studies the Ppapdc1B protein is downregulated in tumor tissue compared to surrounding normal tissue (Figure 7). On the other hand our results are in agreement with what has been observed in HCC, where the PPAPDC1B mRNA levels decrease as the tissue evolves from normal tissue to tumor tissue and in metastasis the Ppapdc1B expression decreases even further (56). The group observing this expression patterns also observed loss of DNA containing this gene and has defined Ppapdc1B as a metastatic tumor suppressor gene (56, 57).

Our results in cell lines, based on the 47 kDa Ppapdc1B and also the 20 kDa Ppapdc1B indicate that the protein expression follows DNA copy number (Figure 5). This is different from what we observe in the tumor samples. It is interesting when considering that cell lines in continuous culture and tumors in the human body do not share a similar environment. In cell culture nutrients and oxygen are found in excess levels so no external influences push the cells through some sort of evolution towards metastatic behavior (8, 9). On the other hand tumors have to struggle to grow and migrate within the human body, like stimulating angiogenesis to ensure sufficient oxygen flow towards the tumor (3). Since the Ppapdc1B protein has been implied as metastatic tumor suppressor gene this high expression in cells in continuous culture might not be surprising. One might speculate that the cells that lack nothing in their environment are not in need for the metastatic features Ppapdc1B inhibits. In addition Ppapdc1B plays a role in renewing membranes by regulating the lipid synthesis

(52, 53). Cell lines in culture are constantly growing fast and are therefore in constant need for renewing membranes. It might therefore be advantage for them to express high levels of Ppapdc1b.

Four breast tumor samples displayed unusually high levels of Ppapdc1B (55 kDa isoform) in our experiments (Figure 6). One of these samples carries gain of the 8p12-p11 amplicon one is defined carrying loss and the other two display normal DNA copy number. This indicates that the protein expression is regulated by more than one mechanism which is also reflected in our results showing no correlation between Ppapdc1B amount and its gene copy number or mRNA amount. Most likely the regulation occurs during the step when the mRNA is translated into protein because DNA copy number and mRNA expression correlate significantly (16). In a way this might explain the contradictory results that have been published about the PPAPDC1B gene in HCC, breast cancer and our study. In all cases, except ours, the gene expression was measured using mRNA, not proteins. Also, the difference in the effect that the Ppapdc1B protein has on cells might be due to the difference of the organs the carcinomas originate in, that is breast and liver.

It should be noticed that the main Ppapdc1B signal observed in all samples does not display at the published molecular weight. All our results are based on the assumption that the antibody is truly detecting Ppapdc1B and not an unrelated protein, after various experiments performed to confirm this. In support of our data the same antibody was used by Dai and colleagues when they were examining Ppapdc1B abundance in HCC. They also observed protein signals at the same molecular weight and their assumptions are the same as ours, that is, the high molecular weight signals represent Ppapdc1B (56). These signals might possibly be dimers or multimers of the protein. According to databases the gene codes for at least three main isoforms that weigh 29, kDa, 24 kDa and 20 kDa, dimers of these proteins would make 40 kDa, 50 kDa and 60 kDa molecules (101, 102). It is possible that these molecules would show signals around 47 – 55 kDa. Ppapdc1B contains few cysteine residues that might possibly form disulfide bridges, yielding dimers or multimers. Disulfide bridges are chemical bonds that can be hard to disrupt (53). However, the western blot protocol includes reducing chemicals to prevent sulfide bridges to form, as described in materials and methods. All proteins should run through the SDS-PAGE displaying their primary structure. In addition, no evidence in the literature shows that Ppadc1b forms multimers.

To verify the signals we see on our western blots are truly Ppapdc1B we could do one of the following: 1) immunoprecipitation where we pull down proteins the antibody recognizes. This is done by adding an antibody, targeting Ppapdc1B, to whole cell lysates which are then loaded on SDS-PAGE for western blots, done with another antibody raised against Ppapdc1B. Even though this would only tell us if the antibody is working properly the antibody should at least not show signals that are background signals. 2) downregulate the gene expression with siRNA and test what signals would disappear after downregulation. 3) overexpress the PPAPDC1B and see what signals appear on a western blot. However this test depends on using cells that process the Ppapdc1B protein normally, even though they are overexpressed from a vector. 4) The last and probably the best option would be to analyze the proteins the antibody is binding, by mass-spectrometry.

5.3. The four candidate genes tested

We used breast cancer cell lines and siRNA to analyze the effect of ERLIN2, LSM1, ASH2L and PROSC on proliferation. The results indicate that none of them do. A gene that can be qualified as an oncogene is able to immortalize and transform normal cells when expressed under the wrong situation that is when the gene is suddenly expressed in excess or is mutated in a way that gives the protein increased function (3, 10). Our experiments aimed to test if our candidate genes were able to influence cell proliferation, a quality that is characterizing for oncogenes. However, even though proteins play an important role in cancer the role might not be exclusive to cancer growth. The gene might be important in cancer progression, maintaining or inducing angiogenesis or the gene might be important for migration of the invasive cancer cells. In this thesis the genes´ ability to stimulate cancer growth was tested; tests regarding other oncogenic abilities of the genes will await further studies.

5.3.1. Erlin2 – is overexpressed in breast tumors but does not affect cell proliferation

Erlin2 and Erlin1 form complexes in the cell that are between one or two MDa large, in the ratio 2:1, respectively (62, 63). Even though the proteins bind together to form this complex the protein expression of either protein is independent of the other. So when Erlin2 is overexpressed Erlin1 is probably stably expressed from chromosome 10 (63). The function of the complex is ubiquitinating IP_3R. The formation of Erlin2 and Erlin1 complex is essential for the function of the proteins. Their function decreases significantly when complex formation is inhibited (62, 64). The complex' function depends on the 2:1 ratio of Erlin2 and Erlin1, for this ratio is the optimal ratio for correct function (62). Overexpression of Erlin1 does not rescue lack of Erlin2, but overexpression of Erlin2 increases ubiquitination of IP_3R (62). In addition, the Erlin2 and Erlin1 complex interacts with other proteins that undergo ERAD pathway for degradation (64). To estimate if the downregulation of Erlin2 is affecting IP_3R ubiquitination, measurements of the IP_3R abundance in the cell lines would have to be done. In our experiments the IP_3R levels were not measured.

Lack of effects from the downregulation on cell proliferation (Figure 10), might be due to the cells lack of dependence on IP_3R. The effects of IP_3R deficiency has been shown in T-cells, cells that belong to the immune system (68). In Jayaraman and Marks experiment the T-cells lacked IP_3R1 completely and IP_3R2 and IP_3R3 partly, which resulted in the cells ability to avoid apoptosis. They tested cells where no IP_3R activation was observed in the T-cells (68). Erlin2 is thought to ubiquitinate all three IP_3R. However only IP_3R1 has been measured in regards of studying Erlin2 function in mammalian cells (64). When IP_3R are activated the Ca^{2+} release from ER and Ca^{2+} introduction into mitochondria are the most important role of IP_3R in inducing apoptosis (119). Our hypothesis was that upregulation of Erlin2 resulted in faster degradation of IP_3R. Where Erlin2 function has been studied, the Erlin2 and Erlin1 complex always bind to IP_3R after activation (59). The overexpression of Erlin2 might not impact the amount of un-activated IP_3R. Therefore measuring IP_3R abundance in cells overexpressing Erlin2 might be interesting, as pointed out before. In addition measuring activated IP_3R and total abundance could give further information on Erlin2 function. Based on our results we cannot exclude that IP_3R or

other substrates of the Erlin2 and Erlin1 complex might play a role in oncogenesis. But the complex interacts with other substrates of the ERAD pathway that have been less studied (64).

It cannot be excluded that the genomic background of the cell lines makes an impact on studies like these. We excluded Erlin2, as a candidate target gene, based on results from two cell lines, CAMA-1 and ZR-75-1, both classify as luminal cells (94-96, 98). To determine whether the observations would be the same for cell lines with different genomic backgrounds one would have to further experiment using different cell lines. To this date the ERLIN2 gene has only been tested in one other cell line that is MCF10A (45). The gene was overexpressed in cells cultured in medium lacking insulin, but MCF10A cells depend on insulin. The overexpression of ERLIN2 resulted in increased focus formation of the cell line, but transformation was not observed for the cells were not able to form colonies in soft agar (45). The cell lines used in our study are isolated from breast tumor (96). They are both immortalized and transformed, unlike MCF10A which is only immortalized (50, 96). The ERLIN2 gene is not sufficient to transform the MCF10A cell line which makes the gene unlikely to drive carcinogenesis (45). In breast cancer cell lines, like ZR-75-1 and CAMA-1 many genomic alterations have accumulated in order to form tumors. Effects from Erlin2 are hard to detect in cell lines when the gene is not even strong enough to transform MCF10A, a cell line that is considered relatively normal (50)

Even though the ERLIN2 gene is highly overexpressed in breast cancer tumors and breast cancer cell lines, the gene does not seem to affect cancer formation or growth (16, 45). The gene might play a role cooperating with other genes on the amplicon or even located on different amplicons. To see the effects of Erlin2 downregulation, knock down of other genes simultaneously might be necessary. In addition our setup might not be able to detect small that effects Erlin2 alone could cause.

5.3.2. LSm1 – our results versus published data

Published data have shown how LSm1 can affect cell proliferation (39, 41, 46). The protein has been downregulated in breast cancer cell lines overexpressing the protein as well as overexpressed in cell lines that are considered normal. Yang and colleagues showed that LSM1 overexpression could induce MCF10A growth in soft agar when the cells were motivated by depriving the cells of insulin in the medium (41, 46). In cooperation with BAG4 and/or C8orf4 LSM1 had the same effects in epidermal growth factor (EGF) deprived medium (41). Taking out factors like insulin and EGF from cell lines that depend on these factors is like a selection force. Thus the gene participated in transforming the immortalized cell line that is used as a representative of normal breast cells in cell-based experiments when cell culture conditions get harder (50). However these alterations in cell culture medium are not always an option in cell culture. The MCF10A cell line, overexpressing LSM1 also grew faster than the normal MCF10A cell line (46). In addition the Streicher group downregulated the gene in a cell line that overexpresses the protein, SUM44. As a result the cell line lost a part of its ability to grow in soft agar (46). In contrast the gene did not affect growth in soft agar when Bernard-Pierrot and her collaborators downregulated the gene in CAMA-1 and ZR-75-1, the same cell lines as

used in our study (39). What they found was that when LSM1 expression was downregulated the cell proliferation decreased and apoptosis was induced (39). The difference in Bernard-Pierrot et al. and Streicher et al. studies is likely due to the differences between cell lines.

It is worrying that our results differ from the results by Bernard-Pierrot´s et al. experiments, since both groups used the same cell lines. As described in materials and methods we tested the cell lines in order to confirm that we were using the same cell lines (purchased from ATCC). Analysis of genetic markers verified that our cell lines had the same DNA profile as listed by the distributor. The cell lines were cultured in virtually the same medium under the same conditions (39). A siRNA molecule was used that targets the same sequence as used by Bernard-Pierrot et al. (39). It should be noted that Bernard-Pierrot et al. measured the mRNA LSM1 level to verify downregulation but we measured the protein level by western blot. We observed few extra signals from the LSm1 antibody, much larger than the expected sizes of the LSm1 protein. These did not decrease in amount despite siRNA treatment. Possible explanation for these extra signals is that the LSm1-7 complexes are not fully degraded apart, so the antibody detects LSm1 in a complex of other LSm proteins. It should also be considered that maybe LSm1 is not degraded continuously in the cells so when we measure the downregulation we are only detecting the decrease in newly synthesized LSm1 (Figure 12). If this is true one might speculate that our downregulation was not as efficient as shown in Bernard-Pierrot's et al. study.

On the other hand few details were different between culture conditions. In our study the cell lines were cultured in DMEM/F12 medium supplemented with 10% FBS and 0.1 nM NEAA or RPMI supplemented with 10% FBS for CAMA-1 and ZR-75-1, respectively. Bernard-Pierrot and her colleagues cultured their cell lines in DMEM supplemented with 10% FBS and 2 mmol/l L-glutamate or RPMI supplemented with 10% FBS, 1mmol/l sodium pyruvate, 4.5 g/l glucose and 10 mmol/l HEPES for CAMA-1 and ZR-75-1, respectively. This difference could possibly have an impact on the results, however because the reagents were purchased from different producers it is possible that Bernard-Pierrot and colleagues were not adding supplements in addition to what is already found in the medium we have purchased. For example, all medium we used contain L-glutamate, a highly stable derivative of L-glutamine molecule and RPMI medium can be bought with high glucose, so adding extra glucose was in our case unnecessary. These are however the only possible explanations we have for the contradictory results in our studies to other published studies, where LSm1 has been pointed out as an oncogene, located within the commonly amplified region 8p12-p11: 1) difference in measurements of downregulation (mRNA vs protein), 2) stability of LSm1-7 complexes (our culture conditions yield more stable complexes) and 3) difference in cell culture conditions. It should be noted that LSm1 did not affect the cells ability to form colonies or grow in soft agar when tested by Bernard-Pierrot et al. (39).

5.3.3. Ash2l – is the down regulation sufficient?

CCDS database describes two isoforms of Ash2l. Isoform 1 weighs 69 kDa and isoform 2 weighs 60 kDa (102). The ensembl database additionally listed five other isoforms that have not received a

CCDS number yet (101). Uniprot database also listed isoform 1 and isoform 2, plus one 56 kDa that matches an isoform from the ensembl database (116).

On our western blots we observed a strong signal of around 80 kDa molecular weight in all cell lines analysed. Ash2l isoform 1 has been shown to run as 80 kDa protein (87). In SUM52 we also observed a signal for a protein weighing around 70 kDa, which could be Ash2l isoform 2. Isoform 2 gives a much weaker signal than isoform 1, which might indicate protein breakdown, saying that this is not isoform 2, but leftovers of isoform1. Since this protein is not observed in other cell lines, or in some cases like CAMA-1 it was only observed sometimes, the signal might be considered a breakdown or background signal.

When Ash2l was downregulated the siRNA treatment in SUM52 didn't manage to target the Ash2l isoform 2 sufficiently. If both signals represent Ash2l the protein expression did not reach levels below D492-3D levels. The knock down was performed using two siRNA molecules targeting ASH2L. The two siRNA molecules were selected based on their ability to target all isoforms of the Ash2l protein, according to the ensembl database (101), and both siRNAs were used in all downregulation experiments. Increasing either of their concentration did not affect the smaller protein detected in SUM52 to the same level as isoform 1. We therefore wonder if the downregulation was not sufficient for Ash2l in SUM52 and therefore no effects were observed (Figure 14). However one might also speculate that the signal is not Ash2l isoform 2 since the siRNA could not affect the protein levels. To be sure the downregulation was sufficient, one might have to increase the siRNA concentration even more or try another siRNA that would be able to downregulate the smaller isoform. The expression of isoform 1 was always measured approximately twofold compared to the expression of isoform 2, both prior to and after knock down. If the expression from isoform 2 is so much lower than isoform 1, it is possible that it does not play a role within the cell, like isoform 1 is thought to do.

Interestingly the Ash2l protein expression in MCF7 was three times higher than Ash2l expression in D492-3D. MCF7 is a breast cancer cell line, but it has been shown to carry a loss on the 8p12-p11 region (Table 5). These results might indicate that increased protein expression from ASH2L is not only controlled by gene copy number. Gene expression is a complicated process that involves both transcription from DNA to mRNA and then translation from mRNA to protein. Various regulations are upheld by the cell to maintain balance. Various known ways to control gene expression are known, such as non-coding RNAs (ncRNAs) (120). Studies have shown that the relationship between DNA levels to mRNA abundance and protein translation are not always linear (10, 121).

ASH2L has not been identified as the amplicons' target gene so far. This gene has been overexpressed in MCF10A cell line, where the cells formed colonies in insulin deprived medium (45). In addition ASH2L overexpression in combination with Ha-RAS or MYC can transform primary rat cells (88). The same group showed that downregulation of ASH2L inhibited cell proliferation in human cancer cells. Further studies are needed to verify these effects found by Lücher-Firzlaff et al (88).

5.3.4. Prosc – the unknown protein

PROSC has never been identified as the putative driver gene of the 8p12-p11 amplicon even though it has been found amplified and overexpressed in few studies (27, 39-41, 43). However it has been tested in cell culture by Yang, where the protein was overexpressed in MCF10A cell line. The overexpression resulted in increased colony formation, when the cells were deprived of insulin in the medium. A slight increase in cell proliferation was observed (45).

When we analyzed the effect of downregulation of the expression of the PROSC gene in breast cancer cell lines we got unexpected results for our control cell line, MCF7. This cell line is thought to carry loss of the 8p12-p11 region (Table 5). Despite the DNA status in this cell, knock down of PROSC resulted in decreased cell proliferation (Figure 16). On the other hand the knock down of PROSC did not affect cell proliferation in the T47D cell line that harbors amplification of the gene, and is also overexpressing the protein (Figure 17). However these results only represent a single experiment, because the downregulation of PROSC was not repeated. Downregulation of PROSC in MCF7 and T47D was tried several times but was only successful this one time represented in this thesis. Due to limited time the experiments were canceled.

To our surprise the Prosc expression in MCF7 was higher than the expression in D492-3D (Figure 17). This is evidence for regulation of the protein expression through different features than amplification. MCF7 and T47D are both classified as luminal cell lines however they differ in genomic background, where MCF7 carries for example normal copy number of the 8p12-p11 amplicon but T47D carries an amplification (95). If the results above are confirmed the reasons why knock down in MCF7 results in decreased cell proliferation but not T47D could be that 1) PROSC is indeed an important gene in cancer progression but the effects vary between cells and their genomic background, that is why the Prosc protein shows as high levels in MCF7 as they do. 2) The downregulation in T47D is not sufficient, so that the effects are not detected. T47D is, after knock down, still expressing more than MCF7 is expressing after treatment with siPROSC. This might reflect that protein levels need to go even lower in T47D. 3) These results actually reflect how working with cell cultures can be hard to manipulate and in this case some unknown factors resulted in decreased cell proliferation for MCF7.

These results do not support the hypothesis that PROSC is the target gene of the amplicon. Even though these results will be verified the genes effects are only observed in a cell line that carries decreased copy number of the gene (Table 5). PROSC might however carry some oncogenic abilities but our results weaken the speculations that it is the target gene of the amplification. Analysis published in 2010 summarizes how target genes of amplifications should only show effects where the gene is overexpressed (10).

5.4. Candidate genes that have not been tested

Our study was limited to four genes, ERLIN2, LSM1, ASH2L and PROSC. The reason these four genes were chosen was due to their high correlation between DNA amplification, mRNA expression

and protein expression in both breast tumors and breast cancer cell lines (16). According to the results of our studies the list of putative target genes of the 8p12-p11 amplification includes seven genes. We have tested the effect of four of these on proliferation and three genes remain untested. These are BRF2, RAB11FIP1 and DDHD2. Each of these three genes is interesting candidate target genes of the 8p12-p11 amplification.

5.4.1. BRF2

TFIIIB-related factor 2 (BRF2) is a gene on 8p12 that codes for a protein that is a subunit of transcription initiating factor III B (TFIIIB). TFIIIB is a complex that binds to promoters of genes coding for none translated RNA molecules and directs RNA polymerase III to its target genes. Disruption in regulation of TFIIIB has been linked to cancer, mostly because of TFIIIB-related factor 1 (BRF1) a similar gene as BRF2. Brf1 regulates transcription of genes such as tRNAs that regulate cell metabolism and proliferation (90, 122). RNA polymerase III is recruited to a certain gene based on the type of TFIIIB complex that binds to the promoter (123). The localization of the promoter contributes to which TFIIIB binds to the sequence, TFIIIB-Brf1 or TFIIIB-Brf2 (90, 123). TFIIIB complex that contains Brf2 stimulates for example transcription of snRNA U6 and 7SK gene (90, 122). Upregulation of Brf2 has been correlated with U6 snRNA transcription rate, when the Brf2 is in excess it results in increased gene expression from Brf2 target genes (122). RNA polymerase III is in fact often linked to cancer formation, but the polymerase regulates gene expression of genes that are highly important for the cells growth (124).

BRF2 has been found amplified and overexpressed in a variety of cancers like gastric and kidney cancer as well as melanoma, where the genes product is found more abundant in cancerous tissue than pairing normal tissue (90). Recent study in lung cancer has also identified amplification and overexpression of BRF2 linked to cell proliferation and focus formation that is in lung squamous cell carcinoma cell lines. The overexpression also led to increased RNA processing mediated through increased transcription of U6 snRNA molecules (125). BRF2 overexpression has been linked to shorter metastatic free survival of breast cancer patients (90). These findings support the fact that BRF2 might drive the amplification of 8p12-p11 chromosomal region in cancers. However, it might be likely to find the BRF2 expression bound to a specific breast cancer subtype (90). Ideas are emerging that the amplicon contains more than one driver gene that might be specific for each subtype, as we will review later in this thesis. To date, Brf2 has been tested once as a putative oncogene in breast cancer, where the gene was overexpressed in MCF10A by Yang and colleagues (45). The overexpression resulted in slightly increased cell proliferation and colony formation, supporting the hypothesis that BRF2 might function as an oncogene (45).

5.4.2. RAB11FIP1

RAB11 family interacting protein 1 (RAB11FIP1) is a member of a family of RAB11 interacting proteins. This protein has also been called Rab coupling protein (RCP) for it was first identified in a yeast two hybrid test where the protein was found to interact with both Rab4 and Rab11 (126). The protein was later found to bind primarily to Rab11 both in vitro and in vivo (91). RAB11FIP1 codes for

eight different isoforms that seem to be differentially located in various subcellular compartments, even though all isoforms are thought to play a role in the plasma membrane recycling system (92). One might speculate that each isoform regulates a subset of proteins for recycling in the cell cycle.

The Rab11fip1-Rab11 complex plays a role in recycling receptors that sit in the plasma membrane, where they play various roles. Rab11fip1 regulates the recycling of receptors located on the surface of the cells which makes the protein a likely candidate for cell differentiation regulation (91, 127). The Rab11fip1-Rab11 complex influences recycling of various proteins, for example Transferrin receptor (TfR), Trans Golgi network protein 2 (TGOLN2), Shiga toxin B (STxB) and Vesicle associated membrane protein 4 (VAMP4) (91, 127). When Rab11fip1 is downregulated the TfR is sent to endo-lysosomes for degradation instead of being recycled and sent back to the plasma membrane (91). The Tgoln2 protein is transferred to trans-Golgi network through transfer that is mediated by Rab11fip1 from early endosomes (EE) to recycling endosomes (RE). But when Rab11fip1 is downregulated Tgoln2 is wrongly localized in various subcellular departments or accumulates in the endosomes (127).

In addition Rab11fip1 has been linked to recycling of integrin proteins. Intergrins are proteins required for cells adhesion to the extra cellular membrane. The integrin proteins are necessary to maintain cell structure. The combination of integrins displayed on the cell surface regulates the cells mobility and structure (128). Rab11fip1 has been shown to induce the recycling of integrin α5β1 and as a result the cells mobility increases and cell migration is promoted (129). If Rab11fip1 is overexpressed in breast cancer the protein might stimulate the tumors´ migration or invasiveness but not necessarily cell growth. In addition a link between Rab11fip1 and epidermal growth factor receptor 1 (EGFR1) has been established, where the receptors recycling is thought to be dependent on α5β1 and its association with Rab11fip1 (129).

Rab11fip1 is often amplified and overexpressed in breast cancer (16, 27, 39-41, 43, 47). In the study of Zhang and colleagues this gene is identified as an oncogene (47). They observed the proteins transformation ability when overexpressed in MCF10A breast cell line. In addition the proteins downregulation in breast cancer cell lines results in decrease of growth in soft agar and cell proliferation (47). The protein also impacts cells migration, this has been shown in MCF10A, MDA-MB-231 and MCF7. In addition when they examine Rab11fip1 mRNA levels in tissue a gradient expression is observed from normal tissue to metastatic tissue, where expression is lowest in normal tissue and increases while the cancer develops (47). These findings are in agreement with the influence of Rab11fip1 on integrin recycling and maintaining of the cell structure as well as the recycling of EGFR1 (47, 129).

5.4.3. DDHD2

DDHD2 codes for a phospholipase $A_{1,}$ an enzyme that catalyses reaction in producing 2-acyl-lysophospholipids and fatty acids (130). Like Rab11fip1 the protein is important for intracellular membrane trafficking towards the plasma membrane (93). Ddhd2 is localized in the Golgi apparatus and in the cytosol (93). The gene is expressed in various tissues in human and the protein is detected

in most cellular compartments except the nucleus and the mitochondria (130). The protein takes part in organizing both Golgi apparatus and endoplasmic reticulum as well. Overexpression of Ddhd2 results in disrupted organization of these organelles of the cell (93, 130). Since the protein was discovered very recently, only few studies have been done to characterize its function in the cell. But Ddhd2 is known to be an enzyme that catalyses the formation of intracellular lipids. These lipids are important for membrane renewal, both in cell divisions, and signals in various pathways (130). Both of these features can have an impact on cancer formation.

DDHD2 has been found amplified and overexpressed in breast cancer, both in our lab as well as others (16, 27, 39, 41, 43). This particular protein has been tested in two separate studies, where the gene is overexpressed (27, 45). Overexpression of the protein resulted in transformation of MCF10A cells when cultured in medium lacking insulin like Yang and colleagues approached it (45). In addition DDHD2 in combination with overexpression of MYC the protein transformed MCF10A in the study carried out by Kwek et al. (27). The findings of Kwek and colleagues are in agreement with the observation that the 8p12-p11 amplicon is often found in combination with amplification of the 8q23-24 chromosomal region, where the MYC gene is localized (15). Knock down of the DDHD2 gene in breast cancer cell lines did not affect cell proliferation or apoptosis. Neither did it affect colony formation or growth in soft agar. This indicates that that DDHD2 alone is not sufficient and it needs cooperation with other factors (39). If the DDHD2 gene was to be downregulated simultaneously as these cooperative factors we might see an impact on breast cancer cell lines growth or even their ability to grow in soft agar.

5.4.4. Regions that lack protein coding genes

Within the minimal amplicon lie not only other genes but also regions that contain no genes at all, “gene-deserts”. It is indeed interesting to view the amplicon, where only a part of the region contains protein coding genes. Even though the region counts 20 protein coding genes, the telomere end of the amplicon only contains ncRNA or no genes (16, 42). Studies on ncRNA molecules for the last decade have shown how non-coding regions play important roles in regulating gene expression and through that cell cycle, cell proliferation and even carcinogenesis (120, 131). On the telomere end of the amplicon few expressed sequence tags (EST) are found. ESTs are mRNA molecules that are expressed from the region but no open reading frame is found so no protein coding gene is defined (132, 133) (www.genome.ucsc.edu, (GRCh37/hg19)). The function of these ESTs has not been identified yet, but they might contain small RNA (sRNA) coding genes or long non-coding RNA (long ncRNA) coding genes. The expression of these ESTs located on the amplified region has not been studied, but if the expression is increased in samples carrying amplification that would be a hint for the possibility of oncogenic function of the RNA molecules. Overexpression of non coding RNAs can alter gene expression and therefore affect the cell cycle. However no data regarding snRNA overexpression are available on BASE. The method that was used for RNA isolation does not detect these small molecules. mRNA was used to measure the expression of genes located on the 8p12-p11 region (16). Estimation of expression of ncRNA from the 8p12-p11 amplicon might reveal interesting information about what this amplification is contributing to tumor growth.

miRNAs regulate gene expression in cells after transcription has occurred, by binding to the 3' end of the mRNA and stabilizes the molecule or inhibits its translation (131). In some cases the mRNA is sent to degradation upon miRNA binding or the mRNA is stored for later use (120, 134). miRNAs have been proven to stimulate and repress cancer growth, based on the molecules´ target genes (131, 135). Within almost all genes located on the minimal amplified region on chromosome 8p12-p11 lies a miRNA target site (132, 133). For example the miR-206 target site is found within the ASH2L gene. ASH2L is often found amplified in luminal breast tumors, which are ER positive (7, 16). miR-206 is a known tumor suppressor miRNA but the molecule inhibits expression of the estrogen receptor α (131). One might speculate that miR-206 binds fewer ER α sequences when ASH2L is amplified due to binding competition to miR-206 target sites. The link between small RNA molecules and the amplification on chromosome 8p12-p11 might be an interesting subject and it is necessary to shed further light on the molecular events occurring in the region.

5.5. Co-amplified regions

Many groups have reported co-amplification of the amplicon on 8p12-p11 along with other chromosome regions, such as 8q24, 11q13, 12q13, 12q15 or 20q13 (15, 22, 27, 28). One of the most common co-amplification and the most studied, is the amplification of 8p12-p11 simultaneously with 11q13. These regions are co-amplified in about 30% of all tumors where 8p12-p11 is found amplified (15). In the tumors where these amplicons are present, mRNA overexpression from both regions is observed (15, 27, 28). The CCND1 gene, on the 11q13 amplicon, is almost always found overexpressed when amplified in breast tumors that harbor both 11q13 amplicon and 8p12-p11 amplicon (15, 28). One study has identified CCND1 to influence the expression of ZNF703, the first protein coding gene found on the telomeric end of the 8p12-p11 amplicon (27). In the same study, CCND1 and ZNF703 were found to, in cooperation, transform MCF10A cell line, when overexpressed (27). Other genes from the 11q13 amplicon have also been listed as possible partners of genes located on 8p12-p11. These genes are for example RPS6KB2 and GAB2, but they have been found significantly amplified in tumors carrying amplification on the 8p12-p11 region (28). Both 8p12-p11 and 11q13 amplicons have been analyzed and mapped into four separate amplified cores, namely A1, A2, A3 and A4 amplicon cores (27, 28, 42). Since both amplicons can be divided into four different cores, one might suggest that each core carries one driver gene that effects tumor development in its specific way. The amplicons' cores can all be found separately or in combination with one another (27, 28, 42). However, systematic analysis on the co-occurrence of these amplicon cores has not been done. Such analysis would have to take place in a large cohort since the possibilities of combinations of all the eight different cores (four cores from each amplicon) are numerous. It needs also to be kept in mind that the 8p12-p11 is not only found co-amplified with 11q13, but also other chromosome regions, as mentioned above, so cooperation between these regions should also be analyzed. This would identify genes within various amplicons that cooperate in tumor progression.

It is highly likely that the amplified genes we find to correlate with increased protein expression require simultaneous overexpression of additional proteins to be able to function as oncogenes. These could be genes located on different chromosomes or within the same chromosomal region. We tested

the latter statement by analyzing the effect of downregulation of more than one gene simultaneously in breast cancer cell lines.

5.5.1. Downregulation of three genes simultaneously

As many groups have pointed out the 8p12-p11 amplicon consists of many genes, each coding for proteins that play various roles in maintaining functioning cells. They play a role in protein trafficking (DDHD2, RAB11FIP1), gene expression regulation (BRF2, LSM1, ASH2L) or protein degradation (ERLIN2) (59, 75, 81, 90, 91, 93). It has not been established clearly enough which genes are targets of the amplification and which genes are passenger genes and if the target is more than one gene. Some groups have identified more than one gene as a candidate target gene, and suggested that the genes need to be overexpressed simultaneously to have an impact on cancer development (27, 41). In our tumor samples ASH2L and LSM1 are along with PROSC most often upregulated simultaneously (16). We therefore tested the effects of downregulating all three genes at the same time (Figure 18). However the results do not support the hypothesis that these three genes are the core stimulation for cancer growth. Downregulation of all three genes simultaneously did not affect cell proliferation in CAMA-1.

The downregulation of three genes at the same time turned out to be rather difficult to perform; the protein expression of CAMA-1 reached the same levels as the protein levels in D492-3D cell line, when three genes were downregulated simultaneously. To be able to see the effects of the downregulation we might have to get the protein expression even lower. One should also consider that the tools we use to measure the knock down might not be accurate enough for us to see the exact amount of the proteins after downregulation. And the measurements of the protein expression in D492-3D are all made from the same protein lysate, so we have not estimated the variation in protein expression from this cell line. So even though we measure the same expression as in D492-3D the protein amount in cells where the protein is downregulated might be higher than normal. In addition to this the downregulation is measured in a pool of cells that were treated with siRNA targeting three genes. Our measurements cannot estimate the portion of cells that do not absorb all siRNA molecules. It should be noticed that some cells might only be downregulating one or two of the three target genes. To be sure the downregulation is sufficient we might have to use other methods to downregulate, like shRNA, which is then expressed in the cell line. This technique would able us to select for cells that are expressing all three shRNAs. What we also benefit from shRNA is the stable downregulation that can continue over longer period of time than siRNA downregulation does. Longer downregulation might reveal effects that only emerge after more than three cell divisions. The cells might be able to overcome downregulation of the genes for a limited time, but after few more cell divisions the effects could emerge. A good example of this is our control gene, PLK1. When the PLK1 gene is downregulated in cells with wild type p53 the PLK1 depletion does not show any effects until the cells have undergone two cell cycles. After two cell cycles the p53 protein is not sufficient to force the cells towards survival (109). In our studies the downregulation was always carried out long enough for the cells to undergo at least two cell cycles. A protein that can be absent for over two cell cycles is unlikely to play an important role in maintaining cancer growth. We also performed a small experiment

where the duration time of the downregulation was measured. The knock down of Erlin2 lasted for at least ten days in CAMA-1. During that time no effects on cell proliferation was obeserved (data not shown). These data indicate that shRNA transfection might not be necessary for longer knock down duration.

In addition the simultaneous downregulation of three genes was only performed in one cell line, this particular cell line is classified as a luminal cell line, that harbors amplification on chromosomal region 8p12-p11 as well as commonly amplified region on chromosome 11q13 (96). The CCND1 gene is located in this amplified region on chromosome 11q in CAMA-1 (96). This gene might be sufficient to rescue the cells when ASH2L, LSM1 and PROSC are all downregulated at the same time. If we would repeat these experiments in another cell line, one which does not harbor amplification on chromosomal region 11q13, we would maybe see some effects of downregulating these three genes.

5.6. Breast cancer suptypes and putative (metastatic) oncogenes

In our search for the amplification target gene we assume the accumulation of its protein gives the tumor cells advantage over those tumor cells not accumulating it. We also assume that the gene in question is not necessarily involved in increased growth it could contribute to invasiveness or migration. Amplification of 8p12-p11 has been linked to bad prognostics and shortened metastatic free survival (22, 23, 28, 40, 42). Many groups have identified a putative target gene of the 8p12-p11 amplificacon but most have identified different genes. The list of identified genes includes at least six genes: ZNF703, LSM1, RAB11FIP1, PPAPDC1B, WHSC1L1 and FGFR1 (39-41, 43, 45-47, 118). In addition some studies have pointed out cooperation between more than one gene within the 8p12-p11 region and also genes located at other chromosomal regions, like 11q (27, 28, 41). As discussed above the driver genes might differ between co-amplification of 8p12-p11 and other various amplified regions. In addition certain regions are co-amplified mainly in certain breast cancer subtypes, for example the co-amplification of 8p12-p11 and 11q13 has been shown to occur primarily in luminal tumors (7).

Emerging knowledge about the variation in breast cancer has opened the possibility of different prospects in studying tumors. Different genomic events that occur in different cells within the same organ can lead to the same phenotype, namely breast cancer. Expression of some genes has been associated with certain sub-populations of breast cancer like FGFR1 which is primarily found affecting lobular breast cancer when overexpressed (117, 118). The FGFR1 gene has been considered as a likely target gene. Studies have indicated that the protein is probably linked to lobular breast cancer and that its oncogenic ability is often connected to other features than cell proliferation, such as migration and invasiveness (110, 118). In addition FGFR1 overexpression has been linked to tumors resistance to endocrine therapy (110). Another gene that has been proposed as the amplicon´s target gene is Zinc finger protein 703 (ZNF703). The effect of that gene seems to be limited to the luminal B breast cancer subtype (40, 43).

5.6.1. ZNF703 as the driver gene of the 8p amplicon in luminal B tumors

Recently two articles were published where ZNF703, a gene localized within the minimal region of the 8p12-p11 amplification, was identified as the amplicons target gene in luminal B breast tumors (40, 43). This gene is amplified and highly overexpressed in tumors that are classified as luminal B subtype (40, 43). When the gene was either overexpressed in MCF7 or downregulated in the MDA-MB-134 and the HCC1500 breast cancer cell lines, effects on cell proliferation were observed. Cells overexpressing ZNF703 have increased cell proliferation. In addition overexpression of the ZNF703 gene resulted in reduced disease free survival (40, 43). This has been shown before for tumors carrying amplification of 8p12-p11 have been linked to worse prognosis (22, 23, 28, 42). In support of these findings another group identified ZNF703 as a target gene by studying the mouse ortholog, *Zeppo1* which is 96% identical to human ZNF703. Overexpression of *Zeppo1* in mouse cell lines resulted in increased cell proliferation, colony formation and disruption in cell structure (136). Interestingly these results were only observed in 3D culture, whereas the protein expression did not affect cells in 2D culture significantly (136). This indicates that the ZNF703 genes main effects are on cancer progression, such as migration and invasion. In fact when expression of E-cadherin was estimated, in a mouse cell line overexpressing *Zeppo1,* Slorach and colleagues found that the cell-cell adhesion molecule was reduced (136). E-cadherin is a protein that connects epithelial cells and helps the cells to maintain polarization, which is important for normal function of cells and the structure of tissue. Disruption of expression of adhesion molecules is evidence for cancer migration and invasion (128, 137). Based on these results one might speculate that the target gene on chromosomal region 8p12-p11 is not necessarily inducing cell proliferation, like we tested in this project. The genes might rather be inducing the cells mobility and encouraging them to migrate into other tissues. Slorach and colleagues also observed reduced lung metastasis in mice when primary tumors where generated from cells with downregulation on *Zeppo1* (136). This indicates that the protein is important in cells ability to metastasize (136).

Znf703 is located in the nucleus, in a protein complex that contains three other proteins (43). Znf703 has been found to increase expression of genes involved in maintaining pluripotency and decreased expression of E-cadherin as well. Znf703 likely regulates the expression of the estrogen receptor (40, 43, 136). In our studies the ZNF703 did not score high enough on correlation studies to be regarded as one of the candidate target genes, based on protein expression of Znf703 (16). These contradicting results might be reflecting the combination of the group of patients used for these studies and different gene products measured. If only mRNA results are considered from studies done in our lab we find that those results are consistent with what both Holland et al. and Sircoulomb et al. find (40, 43).

6. Conclusions

In this study LETM2 and PPAPDC1B were excluded as target genes of the 8p12-p11 amplification and possible effect of four of seven identified candidate target genes on cell proliferation was analyzed. The results indicated that none of these four genes (ERLIN2, LSM1, ASH2L, and PROSC) are likely to affect proliferation. Our studies aim was to identify genes within the 8p12-p11 amplicon that contributes to cell proliferation. However, we cannot exclude the possibility that the experimental setup was not able to detect weak changes in cell proliferation. Minor changes that eventually occur in cell proliferation probably cannot be detected by cell counting. Variations between wells that have undergone different treatment could be similar to variation between wells that have undergone the same treatment.

Emerging evidence is accumulating that support the hypothesis that the 8p12-p11 amplicon does not contain only one driver gene and/or the amplicon needs other chromosomal regions to be amplified and overexpressed to function correctly in regards of helping cancer formation and growth (27, 28, 39-41, 43, 45, 46, 88). Future studies will be needed to identify the driver gene/genes of the 8p12-p11 amplicon and their roles in cancer formation or cancer migration. If the genes are essential for migration like studies have shown, our setup would not detect any changes for it was not intended for the measurement of the effect of downregulation on metastasis (110, 136) .

Future explorations of the 8p12-p11 amplicon in combination with other amplified regions and in regards of different breast cancer subtypes will hopefully bring us closer to understanding the mechanism behind tumor growth and progression. ZNF703 has already been identified as the driver gene for this amplification in the luminal B subtype (40, 43). The amplification is found in other subtypes as well where the target gene has yet to be identified (15, 23).

Of course this thesis does not empty the list of genes that are located on the 8p12-p11 amplicon. First of all the list of candidate genes counted seven genes but in addition the region contains few non-coding RNA genes. The next step would be to continue down the list of candidate genes. For further studies, changes in the experimental setup will be made for identification of the genes effect on metastatic features.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69-90.

2. Oldenburg RA, Meijers-Heijboer H, Cornelisse CJ, Devilee P. Genetic susceptibility for breast cancer: how many more genes to be found? Crit Rev Oncol Hematol. 2007;63(2):125-49.

3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

4. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet. 2008;40(1):17-22.

5. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet. 2003;33 Suppl:238-44.

6. Lux MP, Fasching PA, Beckmann MW. Hereditary breast and ovarian cancer: review and future perspectives. J Mol Med (Berl). 2006;84(1):16-28.

7. Shiu KK, Natrajan R, Geyer FC, Ashworth A, Reis-Filho JS. DNA amplifications in breast cancer: genotypic-phenotypic correlations. Future Oncol. 2010;6(6):967-84.

8. Moelans CB, de Weger RA, Monsuur HN, Vijzelaar R, van Diest PJ. Molecular profiling of invasive breast cancer by multiplex ligation-dependent probe amplification-based copy number analysis of tumor suppressor and oncogenes. Mod Pathol. 2010;23(7):1029-39.

9. Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res. 2010;12(5):R87.

10. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS. A census of amplified and overexpressed human cancer genes. Nat Rev Cancer. 2010;10(1):59-64.

11. Albertson DG. Gene amplification in cancer. Trends Genet. 2006;22(8):447-55.

12. Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR. Genomic instability in breast cancer: pathogenesis and clinical implications. Mol Oncol. 2010;4(3):255-66.

13. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A. 2002;99(20):12963-8.

14. Bautista S, Theillet C. CCND1 and FGFR1 coamplification results in the colocalization of 11q13 and 8p12 sequences in breast tumor nuclei. Genes Chromosomes Cancer. 1998;22(4):268-77.

15. Jönsson G, Staaf J, Vallon-Christersson J, Ringnér 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.

16. Einarsdóttir BÓ. Genomic and expression analysis of genes at the 8p12-p11 amplified region in breast cancer: Identifying potential target genes of the amplification. [Thesis submitted for the Master of Science degree.]2010.

17. Pérez-Losada J, Castellanos-Martín A, Mao JH. Cancer evolution and individual susceptibility. Integr Biol (Camb). 2011;3(4):316-28.

18. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100(14):8418-23.

19. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74.

20. 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.

21. Ciardiello F, McGeady ML, Kim N, Basolo F, Hynes N, Langton BC, et al. Transforming growth factor-alpha expression is enhanced in human mammary epithelial cells transformed by an activated c-Ha-ras protooncogene but not by the c-neu protooncogene, and overexpression of the transforming growth factor-alpha complementary DNA leads to transformation. Cell Growth Differ. 1990;1(9):407-20.

22. 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.

23. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006;10(6):529-41.

24. Staaf J, Jönsson G, Ringnér 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.

25. Kao J, Pollack JR. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer. 2006;45(8):761-9.

26. Krig SR, Miller JK, Frietze S, Beckett LA, Neve RM, Farnham PJ, et al. ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. Oncogene. 2010;29(40):5500-10.

27. Kwek SS, Roy R, Zhou H, Climent J, Martinez-Climent JA, Fridlyand J, et al. Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. Oncogene. 2009;28(17):1892-903.

28. Karlsson E, Waltersson MA, Bostner J, Pérez-Tenorio G, Olsson B, Hallbeck AL, et al. High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1. Genes Chromosomes Cancer. 2011;50(10):775-87.

29. Bocanegra M, Bergamaschi A, Kim YH, Miller MA, Rajput AB, Kao J, et al. Focal amplification and oncogene dependency of GAB2 in breast cancer. Oncogene. 2010;29(5):774-9.

30. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer Res. 2011;71(1):78-86.

31. Colombo PE, Milanezi F, Weigelt B, Reis-Filho JS. Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. Breast Cancer Res. 2011;13(3):212.

32. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med. 2010;7(5):e1000279.

33. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res. 2010;16(21):5222-32.

34. Subik K, Lee JF, Baxter L, Strzepek T, Costello D, Crowley P, et al. The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. Breast Cancer (Auckl). 2010;4:35-41.

35. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009;360(8):790-800.

36. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011;5(1):5-23.

37. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol. 2010;28(20):3271-7.

38. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.

39. Bernard-Pierrot I, Gruel N, Stransky N, Vincent-Salomon A, Reyal F, Raynal V, et al. Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer Res. 2008;68(17):7165-75.

40. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, et al. ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. EMBO Mol Med. 2011;3(3):167-80.

41. Yang ZQ, Streicher KL, Ray ME, Abrams J, Ethier SP. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer Res. 2006;66(24):11632-43.

42. Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rougé C, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res. 2005;3(12):655-67.

43. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. EMBO Mol Med. 2011;3(3):153-66.

44. Cuny M, Kramar A, Courjal F, Johannsdottir V, Iacopetta B, Fontaine H, et al. Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer Res. 2000;60(4):1077-83.

45. Yang ZQ, Liu G, Bollig-Fischer A, Giroux CN, Ethier SP. Transforming properties of 8p11-12 amplified genes in human breast cancer. Cancer Res. 2010;70(21):8487-97.

46. Streicher KL, Yang ZQ, Draghici S, Ethier SP. Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon. Oncogene. 2007;26(14):2104-14.

47. Zhang J, Liu X, Datta A, Govindarajan K, Tam WL, Han J, et al. RCP is a human breast cancer-promoting gene with Ras-activating function. J Clin Invest. 2009;119(8):2171-83.

48. Tamai S, Iida H, Yokota S, Sayano T, Kiguchiya S, Ishihara N, et al. Characterization of the mitochondrial protein LETM1, which maintains the mitochondrial tubular shapes and interacts with the AAA-ATPase BCS1L. J Cell Sci. 2008;121(Pt 15):2588-600.

49. Waldeck-Weiermair M, Jean-Quartier C, Rost R, Khan MJ, Vishnu N, Bondarenko AI, et al. Leucine Zipper EF Hand-containing Transmembrane Protein 1 (Letm1) and Uncoupling Proteins 2 and 3 (UCP2/3) Contribute to Two Distinct Mitochondrial Ca2+ Uptake Pathways. J Biol Chem. 2011;286(32):28444-55.

50. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. 1990;50(18):6075-86.

51. Carman GM, Han GS. Roles of phosphatidate phosphatase enzymes in lipid metabolism. Trends Biochem Sci. 2006;31(12):694-9.

52. Carman GM, Han GS. Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. J Biol Chem. 2009;284(5):2593-7.

53. Takeuchi M, Harigai M, Momohara S, Ball E, Abe J, Furuichi K, et al. Cloning and characterization of DPPL1 and DPPL2, representatives of a novel type of mammalian phosphatidate phosphatase. Gene. 2007;399(2):174-80.

54. Sciorra VA, Morris AJ. Roles for lipid phosphate phosphatases in regulation of cellular signaling. Biochim Biophys Acta. 2002;1582(1-3):45-51.

55. Jamal Z, Martin A, Gomez-Muñoz A, Brindley DN. Plasma membrane fractions from rat liver contain a phosphatidate phosphohydrolase distinct from that in the endoplasmic reticulum and cytosol. J Biol Chem. 1991;266(5):2988-96.

56. Dai C, Dong QZ, Ren N, Zhu JJ, Zhou HJ, Sun HJ, et al. Downregulation of HTPAP transcript variant 1 correlates with tumor metastasis and poor survival in patients with hepatocellular carcinoma. Cancer Sci. 2011;102(3):583-90.

57. Wu X, Jia HL, Wang YF, Ren N, Ye QH, Sun HC, et al. HTPAP gene on chromosome 8p is a candidate metastasis suppressor for human hepatocellular carcinoma. Oncogene. 2006;25(12):1832-40.

58. Browman DT, Resek ME, Zajchowski LD, Robbins SM. Erlin-1 and erlin-2 are novel members of the prohibitin family of proteins that define lipid-raft-like domains of the ER. J Cell Sci. 2006;119(Pt 15):3149-60.

59. Pearce MM, Wang Y, Kelley GG, Wojcikiewicz RJ. SPFH2 mediates the endoplasmic reticulum-associated degradation of inositol 1,4,5-trisphosphate receptors and other substrates in mammalian cells. J Biol Chem. 2007;282(28):20104-15.

60. Sprenger RR, Fontijn RD, van Marle J, Pannekoek H, Horrevoets AJ. Spatial segregation of transport and signalling functions between human endothelial caveolae and lipid raft proteomes. Biochem J. 2006;400(3):401-10.

61. Browman DT, Hoegg MB, Robbins SM. The SPFH domain-containing proteins: more than lipid raft markers. Trends Cell Biol. 2007;17(8):394-402.

62. Pearce MM, Wormer DB, Wilkens S, Wojcikiewicz RJ. An endoplasmic reticulum (ER) membrane complex composed of SPFH1 and SPFH2 mediates the ER-associated degradation of inositol 1,4,5-trisphosphate receptors. J Biol Chem. 2009;284(16):10433-45.

63. Hoegg MB, Browman DT, Resek ME, Robbins SM. Distinct regions within the erlins are required for oligomerization and association with high molecular weight complexes. J Biol Chem. 2009;284(12):7766-76.

64. Wang Y, Pearce MM, Sliter DA, Olzmann JA, Christianson JC, Kopito RR, et al. SPFH1 and SPFH2 mediate the ubiquitination and degradation of inositol 1,4,5-trisphosphate receptors in muscarinic receptor-expressing HeLa cells. Biochim Biophys Acta. 2009;1793(11):1710-8.

65. Patterson RL, Boehning D, Snyder SH. Inositol 1,4,5-trisphosphate receptors as signal integrators. Annu Rev Biochem. 2004;73:437-65.

66. Vanderheyden V, Devogelaere B, Missiaen L, De Smedt H, Bultynck G, Parys JB. Regulation of inositol 1,4,5-trisphosphate-induced Ca2+ release by reversible phosphorylation and dephosphorylation. Biochim Biophys Acta. 2009;1793(6):959-70.

67. Decuypere JP, Monaco G, Bultynck G, Missiaen L, De Smedt H, Parys JB. The IP(3) receptor-mitochondria connection in apoptosis and autophagy. Biochim Biophys Acta. 2011;1813(5):1003-13.

68. Jayaraman T, Marks AR. T cells deficient in inositol 1,4,5-trisphosphate receptor are resistant to apoptosis. Mol Cell Biol. 1997;17(6):3005-12.

69. Fischer S, Benz J, Späth B, Maier LK, Straub J, Granzow M, et al. The archaeal Lsm protein binds to small RNAs. J Biol Chem. 2010;285(45):34429-38.

70. Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Séraphin B. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J. 1999;18(12):3451-62.

71. Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. Yeast Sm-like proteins function in mRNA decapping and decay. Nature. 2000;404(6777):515-8.

72. Thore S, Mayer C, Sauter C, Weeks S, Suck D. Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J Biol Chem. 2003;278(2):1239-47.

73. Ingelfinger D, Arndt-Jovin DJ, Lührmann R, Achsel T. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. RNA. 2002;8(12):1489-501.

74. Totaro A, Renzi F, La Fata G, Mattioli C, Raabe M, Urlaub H, et al. The human Pat1b protein: a novel mRNA deadenylation factor identified by a new immunoprecipitation technique. Nucleic Acids Res. 2011;39(2):635-47.

75. Mullen TE, Marzluff WF. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev. 2008;22(1):50-65.

76. Ozgur S, Chekulaeva M, Stoecklin G. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. Mol Cell Biol. 2010;30(17):4308-23.

77. Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B. A Sm-like protein complex that participates in mRNA degradation. EMBO J. 2000;19(7):1661-71.

78. Watson PM, Miller SW, Fraig M, Cole DJ, Watson DK, Boylan AM. CaSm (LSm-1) overexpression in lung cancer and mesothelioma is required for transformed phenotypes. Am J Respir Cell Mol Biol. 2008;38(6):671-8.

79. Shearn A, Rice T, Garen A, Gehring W. Imaginal disc abnormalities in lethal mutants of Drosophila. Proc Natl Acad Sci U S A. 1971;68(10):2594-8.

80. Ikegawa S, Isomura M, Koshizuka Y, Nakamura Y. Cloning and characterization of ASH2L and Ash2l, human and mouse homologs of the Drosophila ash2 gene. Cytogenet Cell Genet. 1999;84(3-4):167-72.

81. Patel A, Vought VE, Dharmarajan V, Cosgrove MS. A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem. 2011;286(5):3359-69.

82. Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol. 2006;13(8):713-9.

83. Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol. 2006;13(9):852-4.

84. Cao F, Chen Y, Cierpicki T, Liu Y, Basrur V, Lei M, et al. An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. PLoS One. 2010;5(11):e14102.

85. Justin N, De Marco V, Aasland R, Gamblin SJ. Reading, writing and editing methylated lysines on histone tails: new insights from recent structural studies. Curr Opin Struct Biol. 2010;20(6):730-8.

86. Tan CC, Sindhu KV, Li S, Nishio H, Stoller JZ, Oishi K, et al. Transcription factor Ap2δ associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription. Proc Natl Acad Sci U S A. 2008;105(21):7472-7.

87. Xu Z, Gong Q, Xia B, Groves B, Zimmermann M, Mugler C, et al. A role of histone H3 lysine 4 methyltransferase components in endosomal trafficking. J Cell Biol. 2009;186(3):343-53.

88. Lüscher-Firzlaff J, Gawlista I, Vervoorts J, Kapelle K, Braunschweig T, Walsemann G, et al. The human trithorax protein hASH2 functions as an oncoprotein. Cancer Res. 2008;68(3):749-58.

89. Ikegawa S, Isomura M, Koshizuka Y, Nakamura Y. Cloning and characterization of human and mouse PROSC (proline synthetase co-transcribed) genes. J Hum Genet. 1999;44(5):337-42.

90. Cabarcas S, Schramm L. RNA polymerase III transcription in cancer: the BRF2 connection. Mol Cancer. 2011;10:47.

91. Peden AA, Schonteich E, Chun J, Junutula JR, Scheller RH, Prekeris R. The RCP-Rab11 complex regulates endocytic protein sorting. Mol Biol Cell. 2004;15(8):3530-41.

92. Jin M, Goldenring JR. The Rab11-FIP1/RCP gene codes for multiple protein transcripts related to the plasma membrane recycling system. Biochim Biophys Acta. 2006;1759(6):281-95.

93. Sato S, Inoue H, Kogure T, Tagaya M, Tani K. Golgi-localized KIAA0725p regulates membrane trafficking from the Golgi apparatus to the plasma membrane in mammalian cells. FEBS Lett. 2010;584(21):4389-95.

94. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One. 2009;4(7):e6146.

95. Mackay A, Tamber N, Fenwick K, Iravani M, Grigoriadis A, Dexter T, et al. A high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines. Breast Cancer Res Treat. 2009;118(3):481-98.

96. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006;10(6):515-27.

97. Pole JC, Courtay-Cahen C, Garcia MJ, Blood KA, Cooke SL, Alsop AE, et al. High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. Oncogene. 2006;25(41):5693-706.

98. Jönsson G, Staaf J, Olsson E, Heidenblad M, Vallon-Christersson J, Osoegawa K, et al. High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. Genes Chromosomes Cancer. 2007;46(6):543-58.

99. Cowell JK, LaDuca J, Rossi MR, Burkhardt T, Nowak NJ, Matsui S. Molecular characterization of the t(3;9) associated with immortalization in the MCF10A cell line. Cancer Genet Cytogenet. 2005;163(1):23-9.

100. Gudjonsson T, Villadsen R, Nielsen HL, Rønnov-Jessen L, Bissell MJ, Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. Genes Dev. 2002;16(6):693-706.

101. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, et al. Ensembl 2011. Nucleic Acids Res. 2011;39(Database issue):D800-6.

102. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. 2009;19(7):1316-23.

103. Lambros MB, Natrajan R, Geyer FC, Lopez-Garcia MA, Dedes KJ, Savage K, et al. PPM1D gene amplification and overexpression in breast cancer: a qRT-PCR and chromogenic in situ hybridization study. Mod Pathol. 2010;23(10):1334-45.

104. Pärssinen J, Alarmo EL, Karhu R, Kallioniemi A. PPM1D silencing by RNA interference inhibits proliferation and induces apoptosis in breast cancer cell lines with wild-type p53. Cancer Genet Cytogenet. 2008;182(1):33-9.

105. Fiscella M, Zhang H, Fan S, Sakaguchi K, Shen S, Mercer WE, et al. Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner. Proc Natl Acad Sci U S A. 1997;94(12):6048-53.

106. Rayter S, Elliott R, Travers J, Rowlands MG, Richardson TB, Boxall K, et al. A chemical inhibitor of PPM1D that selectively kills cells overexpressing PPM1D. Oncogene. 2008;27(8):1036-44.

107. Yu E, Ahn YS, Jang SJ, Kim MJ, Yoon HS, Gong G, et al. Overexpression of the wip1 gene abrogates the p38 MAPK/p53/Wip1 pathway and silences p16 expression in human breast cancers. Breast Cancer Res Treat. 2007;101(3):269-78.

108. Petronczki M, Lénárt P, Peters JM. Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1. Dev Cell. 2008;14(5):646-59.

109. Guan R, Tapang P, Leverson JD, Albert D, Giranda VL, Luo Y. Small interfering RNA-mediated Polo-like kinase 1 depletion preferentially reduces the survival of p53-defective, oncogenic transformed cells and inhibits tumor growth in animals. Cancer Res. 2005;65(7):2698-704.

110. Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, Lopez-Garcia MA, et al. FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. Cancer Res. 2010;70(5):2085-94.

111. Tsuji K, Kawauchi S, Saito S, Furuya T, Ikemoto K, Nakao M, et al. Breast cancer cell lines carry cell line-specific genomic alterations that are distinct from aberrations in breast cancer tissues: comparison of the CGH profiles between cancer cell lines and primary cancer tissues. BMC Cancer. 2010;10:15.

112. Loi S. Molecular analysis of hormone receptor positive (luminal) breast cancers: what have we learnt? Eur J Cancer. 2008;44(18):2813-8.

113. Tharun S. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? RNA Biol. 2009;6(3):228-32.

114. Ingelfinger D, Arndt-Jovin DJ, Luhrmann R, Achsel T. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. RNA. 2002;8(12):1489-501. Epub 2003/01/08.

115. Ozgur S, Chekulaeva M, Stoecklin G. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. Mol Cell Biol. 2010;30(17):4308-23. Epub 2010/06/30.

116. Consortium U. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 2011;39(Database issue):D214-9.

117. Reis-Filho JS, Simpson PT, Turner NC, Lambros MB, Jones C, Mackay A, et al. FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. Clin Cancer Res. 2006;12(22):6652-62.

118. Xian W, Pappas L, Pandya D, Selfors LM, Derksen PW, de Bruin M, et al. Fibroblast growth factor receptor 1-transformed mammary epithelial cells are dependent on RSK activity for growth and survival. Cancer Res. 2009;69(6):2244-51.

119. Joseph SK, Hajnóczky G. IP3 receptors in cell survival and apoptosis: Ca2+ release and beyond. Apoptosis. 2007;12(5):951-68.

120. Turner AM, Morris KV. Controlling transcription with noncoding RNAs in mammalian cells. Biotechniques. 2010;48(6):ix-xvi.

121. Srivastava N, Manvati S, Srivastava A, Pal R, Kalaiarasan P, Chattopadhyay S, et al. miR-24-2 controls H2AFX expression regardless of gene copy number alteration and induces apoptosis by targeting antiapoptotic gene BCL-2: a potential for therapeutic intervention. Breast Cancer Res. 2011;13(2):R39.

122. Cabarcas S, Jacob J, Veras I, Schramm L. Differential expression of the TFIIIB subunits Brf1 and Brf2 in cancer cells. BMC Mol Biol. 2008;9:74.

123. Schramm L, Pendergrast PS, Sun Y, Hernandez N. Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. Genes Dev. 2000;14(20):2650-63.

124. White RJ. RNA polymerase III transcription and cancer. Oncogene. 2004;23(18):3208-16.

125. Lockwood WW, Chari R, Coe BP, Thu KL, Garnis C, Malloff CA, et al. Integrative genomic analyses identify BRF2 as a novel lineage-specific oncogene in lung squamous cell carcinoma. PLoS Med. 2010;7(7):e1000315.

126. Lindsay AJ, Hendrick AG, Cantalupo G, Senic-Matuglia F, Goud B, Bucci C, et al. Rab coupling protein (RCP), a novel Rab4 and Rab11 effector protein. J Biol Chem. 2002;277(14):12190-9.

127. Jing J, Junutula JR, Wu C, Burden J, Matern H, Peden AA, et al. FIP1/RCP binding to Golgin-97 regulates retrograde transport from recycling endosomes to the trans-Golgi network. Mol Biol Cell. 2010;21(17):3041-53.

128. Hynes RO. Integrins: bidirectional, allosteric signaling machines. Cell. 2002;110(6):673-87.

129. Caswell PT, Chan M, Lindsay AJ, McCaffrey MW, Boettiger D, Norman JC. Rab-coupling protein coordinates recycling of α5β1 integrin and EGFR1 to promote cell migration in 3D microenvironments. J Cell Biol. 2008;183(1):143-55.

130. Nakajima K, Sonoda H, Mizoguchi T, Aoki J, Arai H, Nagahama M, et al. A novel phospholipase A1 with sequence homology to a mammalian Sec23p-interacting protein, p125. J Biol Chem. 2002;277(13):11329-35.

131. Yu Z, Baserga R, Chen L, Wang C, Lisanti MP, Pestell RG. microRNA, cell cycle, and human breast cancer. Am J Pathol. 2010;176(3):1058-64.

132. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

133. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 2011;39(Database issue):D876-82.

134. Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. Nat Rev Mol Cell Biol. 2007;8(1):23-36.

135. O'Day E, Lal A. MicroRNAs and their target gene networks in breast cancer. Breast Cancer Res. 2010;12(2):201.

136. Slorach EM, Chou J, Werb Z. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes Dev. 2011;25(5):471-84.

137. Ishiyama N, Lee SH, Liu S, Li GY, Smith MJ, Reichardt LF, et al. Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. Cell. 2010;141(1):117-28.