Mögnun og genatjáning á litningasvæði 8p12-p11 í brjóstaæxlum

Líkleg markgen mögnunarinnar tilgreind

Berglind Ósk Einarsdóttir

Ritgerð til meistaragráðu í Líf- og læknavísindum

Leiðbeinandi: Inga Reynisdóttir

Umsjónarkennari: Bjarni A. Agnarsson

Meistaranámsnefnd: Inga Reynisdóttir, Bjarni A. Agnarsson

og Óskar Þór Jóhannsson

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2010

Genomic and expression analysis of genes at the 8p12-p11 amplified region in breast cancer

Identifying potential target genes of the amplified region

Berglind Ósk Einarsdóttir

A thesis submitted for the Master of Science degree

Supervisor: Inga Reynisdóttir

Supervisory teahcer: Bjarni A. Agnarsson

Masters committee: Inga Reynisdóttir, Bjarni A. Agnarsson

and Óskar Þór Jóhannsson

University of Iceland, Faculty of Medicine

School of health science

June 2010

Prentun: Háskólaprent

Staður, Ísland 2010

ÁGRIP

Brjósta krabbamein er algengasta krabbameinið í konum í vestur Evrópu. Það er misleitur sjúkdómur sem orsakast aðallega af breytingum í erfðaefni og umhverfi frumanna. Breytingar í erfðaefni fruma geta breytt eðlilegum frumum í æxlismyndandi krabbameinsfrumur. Magnanir á litningasvæðum sem innihalda æxlisgen gefa frumum yfirburði sem veldur vali fyrir þeim. Mögnun á litningasvæði 8p12-p11 er þekkt breyting í brjósta æxlum sem hefur verið tengd verri horfum. Nokkrir rannsóknarhópar hafa rannsakað svæðið með tilliti til þess að tilgreina markgen mögnunarinnar. Þrátt fyrir það hefur markgenið ekki enn verið tilgreint. Markmið þessarar rannsóknar var að kortleggja 8p12-p11 mögnunarsvæðið í brjósta æxlum og tilgreina möguleg markgen. Fjögur skilyrði voru sett sem gen þurftu að standast til að vera álitin möguleg markgen mögnunarinnar. Í brjóstaæxlum þurftu mögnuð gen að vera yfirtjáð. Þau gen sem voru bæði mögnuð og yfirtjáð þurftu að sýna fylgni milli prótein tjáningar og bæði DNA eintaka fjölda og mRNA eintakafjölda. Einnig var sett það skilyrði að prótein tjáning gensins væri hærri í frumulínum með mögnun á því svæði sem genið er staðsett en í frumulínum og þekjuvefsfrumum úr brjóstaminnkunar aðgerðum sem ekki eru með mögnun á svæðinu. Nýttar voru upplýsingar um DNA eintaka fjölda í brjóstaæxlis sýnum sem aðgengilegar voru í gegnum gagnabanka, til að kortleggja 8p12-p11 litningasvæðið. Minnsta sameiginlega svæði mögnunar var skilgreint sem 1,6 Mb svæði sem inniheldur 20 prótein kóðandi gen. Upplýsingar um mRNA eintaka

fjölda frá sama gagnabanka voru nýttar til að athuga fylgni milli DNA eintaka fjölda og mRNA eintaka fjölda. Þrettán gen sýndu fylgni þar á milli og var prótein tjáning þeirra könnuð. Prótein tjáning genanna þrettán gaf til kynna sjö gen sem líkleg markgen mögnunarsvæðisins 8p12-p11 í brjóstaæxlum.

ABSTRACT

Breast cancer is the most frequent cancer in women in Western Europe. It is a heterogeneous disease and mainly the consequence of alterations in environmental, genetic and epigenetic factors. Alterations in the genetic material of cells can drive the progressive transformation of normal cells into highly malignant tumour cells. Amplification at a chromosomal region containing oncogenes(s) gives cells strong selective advantage. The amplification at 8p12-p11 is a well known alteration in breast cancer and has been linked to poor prognosis. Several research groups have analysed the region with the aim of identifying the target gene(s) that drives the amplification. Despite that, the definite target gene(s) has not been identified yet. The aims of this study were to map the 8p12-p11 amplified region in breast tumours and analyse protein expression of genes at the region to identify potential target genes. Four criteria were set for a gene to be considered a likely target gene. In breast tumours, the gene had to be overexpressed when amplified, the overexpressed gene had to show correlation between protein expression and DNA copy number and protein expression of the genes had to correlate with mRNA expression. Also, the protein encoded by the gene had to be expressed at a higher level in at least 3 of 4 cell lines with amplification of the gene than in cell lines and primary cells without the amplification. Using array-CGH data from available databank including the DNA copy number in breast samples, the minimal region of amplification was mapped as a 1.6 Mb region where 20 known protein coding genes

reside. Using array-GEx results from the same databank, correlation between the DNA copy number and mRNA level revealed thirteen genes as overexpressed when amplified. Protein expression analysis in breast tumours, breast cell lines and primary epithelial cells indicated seven of these genes as the most likely candidate target genes of the 8p12-p11 amplified region based on that they met all the criteria.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors Inga Reynisdóttir and Rósa Björk Barkardóttir for giving me the opportunity to work on this project. Their excellent guidance and expertise were of great support throughout this project. I would like to sincerely thank my co-workers at the Molecular Pathology Unit Guðrún Jóhannesdóttir, Þórgunnur E. Pétursdóttir, Edda Bjarnadóttir, Valgarður Egilsson, Haukur Gunnarsson and Aðalgeir Arason for their great motivation and stimulating discussions. Special thanks go to Haukur and Aðalgeir for their technical advice.

I would like to thank my M.S committee, Bjarni A. Agnarsson and Óskar Þór Jóhannsson for taking the time to contribute to this project and their helpful comments. I would like to thank Bjarni A. Agnarsson especially for his pathological part and Óskar Þór Jóhannsson for his clinical part of this study.

I would like to thank the Stem Cell Research Unit team for all their technical support. Their excellent knowledge and good spirit were of great advantage to me. Especially I would like to thank Þórarinn Guðjónsson for the contribution of cell lines.

Last but not least I would like to express my deepest gratitude to my family and friends, for their unfailing support and encouragement. My parents and parents-in-law get special thanks for all their help. Also I would like to thank my fiancé Siggeir Fannar Brynjólfsson from the bottom of my heart for all his motivation and endless support,

especially for putting up with me in the process of writing this thesis. This thesis is dedicated to my lovely daughter Krista Ýr Siggeirsdóttir for being the joy of my life.

This study was supported by funding from The Icelandic Research Fund, The Research Fund of Landspítali-University Hospital and The Icelandic association “Walking for Breast Cancer”

.

TABLE OF CONTENTS

FIGURES AND TABLES

ABBREVIATIONS

ADRB3	adrenergic receptor, beta 3
ASH2L	ash2 (absent, small, or homeotic)-like (Drosophila)
BAC	Bacterial artificial chromosome
BAG4	BCL2-associated athanogene 4
BASE	Bioarray Software Environment
BCA	Bicinchroninic acid assay
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
BRF2	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
CCND1	Cyclin D1
CGH	(array) Comparative genomic hybridization
DDHD2	DDHD domain containing 2
DNA	Deoxyribonucleic acid
EGF	Epidermal growth factor
EIF1EBP1	eukaryotic translation initiation factor 4E binding protein 1
ER	Estrogen receptor
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GEx	(array) Genomic expression
GOT1L1	glutamic-oxaloacetic transaminase 1-like 1
GPR124	G protein-coupled receptor 124
HSR	Homogenously staining region
kDa	Kilo dalton
LEP	Luminal epithelial cells
LETM2	leucine zipper-EF-hand containing transmembrane protein 2
LOH	Loss of heterozygosity

LSM1	LSM1 homolog, U6 small nuclear RNA associated (S. cerevisiae)
Mb	Mega base
MEP	myo epithelial cells
mRNA	messenger ribonucleic acid
NSD3/*WHSC1L1*	Wolf-Hirschhorn syndrome candidate 1-like 1
PPAPDC1B	phosphatidic acid phosphatase type 2 domain containing 1B
PR	Progesterone receptor
PROSC	proline synthetase co-transcribed
PTEN	Phosphatase and tensin homolog
RAB11FIP1	RAB11 family interacting protein 1 (class I)
RNA	Ribonucleic acid
rRNA	ribosomal RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SNR	Signal to noise ratio
SPFH2/ERLIN2	ER lipid raft association 2
STAR	Steroidogenic acute regulatory protein
TACC1	Transforming, acidic coiled-coil containing protein 1
TNBT	Triple negative breast tumour
TP53	Tumour protein p53
ZNF703	Zinc finger protein 703

Introduction

Epidemiology of breast cancer

Breast cancer is the most frequent cancer in women accounting for 23% of all cancer types. Age standardised incidence and mortality rate in Western Europe is 84.6/100.000 and 22.3/100.000, respectively. The incidence rate is highest in the developed areas; Northern and Southern America, Western and Northern Europe and Australia/New Zealand (Figure 1) (Parkin et al., 2005). Difference in cultural-, dietary- and environmental factors could account for some of the variation.

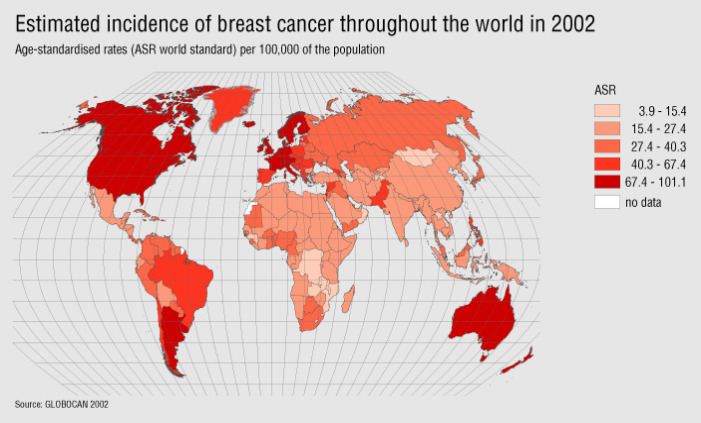


Figure 1. Global breast cancer incidence rates. Rates are age standardized per 100.000 of the population. Rate level is colour coded, see scale on the right side. http://www.euphix.org/object_document/o5614n27171.html (EUPHIX, Bilthoven: RIVM, version1.10, 22 October 2009 (Pisani)).

Survival rate is fairly high for breast cancer, an estimated 74% in Western Europe, age adjusted (Parkin et al., 2005). The incidence rate has been increasing while mortality rate has declined. Better

diagnostic tools along with diagnosis at earlier stage of the disease and better treatment resources are believed to account for the change.

Many factors have been linked to increased risk of breast cancer. Endogenous estrogen exposure such as pregnancy, early menarche, late menopause, and obesity along with exogenous estrogens such as oral contraceptives and hormonal replacement therapies have been associated with higher probability of breast cancer (Verheul et al., 2000). Higher breast cancer risk is observed in monozygotic twins if one of them develops the disease than in dizygotic twins, indicating that genetic factors play a role in cancer development (Peto & Mack, 2000).

Familial breast cancer

Increased risk of tumour formation is found in relatives of breast cancer patients. Around 10-15% of breast cancer patients come from families with high risk of breast cancer and in about half of them the disease can be explained by segregation of a high penetrance breast cancer gene (reviewed in (Lux et al., 2006)). Mutations in the DNA repair genes *BRCA1* and *BRCA2* are well defined high penetrance cancer genes and among the most studied. Mutations within *BRCA1* and *BRCA2* confer a high risk of breast cancer with 10-20 fold increase of relative risk compared to the general population (reviewed in (Stratton & Rahman, 2008)). In Iceland, a mutation within the *BRCA1* gene is very rare but one recurrent *BRCA2* mutation (999del5) is relatively frequent and explains around 40% of Icelandic hereditary breast cancer (Gudmundsson et al., 1996; Thorlacius et al., 1996). Breast cancer patients who have inherited the tendency of tumourigenesis but do not have mutations in *BRCA1* or

BRCA2 are categorised as BRCAX. A common genetic alteration has not yet been identified in the BRCAX patient group with such a high frequency as mutations in *BRCA1* and *BRCA2*. Other breast cancer susceptibility genes have been identified at a lower frequency. A germline mutation in *TP53* is found in over half of patients with Li-Fraumeni syndrome (LFS). In addition to breast cancer, three other cancers are observed in LFS, they are sarcoma, brain tumour and adrenocortical carcinoma. A germline mutation in the *PTEN* gene accounts for most cases of Cowden disease (CS) which is a rare familiar trait characterized by hamartomas and predispositions to breast cancer among other malignancies. Mutations in *STK11* (Peutz-Jeghers syndrome) and *CDH1* are also associated with elevated risks of breast cancer (reviewed in (Stratton & Rahman, 2008; Blanco et al., 2009)).

Low penetrance breast cancer susceptibility alleles have been identified in recent years. Some of these breast cancer risk alleles are within regions that harbour known protein coding genes, for example *FGFR2* (encoding fibroblast growth factor receptor 2) and *TOX3* (encoding a protein that might act as a transcription factor) (reviewed in (Stratton & Rahman, 2008)). However, some low penetrance risk alleles are within regions where no known genes reside.

Emphasis has also been placed on classification of breast cancer subtypes and technological advances of detecting somatic genetic alterations that aid in the identification of disease pathogenesis, as will be discussed in more detail in the following sections.

Breast cancer pathology

Breast cancer arising from epithelial cells is called breast carcinoma, the two epithelial cell types primarily affected by breast cancer are ductal and lobular cells. Invasive ductal carcinoma (IDC) is the most frequent form of carcinoma accounting for around 80% of all cases while invasive lobular carcinoma (ILC) comprises around 10% of all cases (Bharat et al., 2009).

Pathological factors have long been used for prognostic predictions and adjuvant therapy recommendations for breast cancer patients. Tumour grade, lymph node status, endocrine receptor status and epidermal growth factor receptor 2 (ERBB2/HER2) statuses are the main decision factors used. However, these factors can in some cases be misleading and patients can be overtreated and undertreated as a result.

Over the past decade microarray-Gene Expression (array-GEx) technology has been in development for more detailed classification of breast tumours into clinically relevant subgroups. Hierarchical clustering using gene expression data has been shown to organize tumour samples into groups with distinct gene expression patterns and significant different clinical outcomes (Perou et al., 2000; Sorlie et al., 2001; Sorlie et al., 2003; Hu et al., 2006). According to the microarray data malignant breast tumours can be classified into 5 subgroups based on their gene expression profile; luminal A, luminal B, basal-like, ERBB2-overexpressing and normal breast tissue-like subgroups. Luminal A tumours are estrogen receptor (ER) positive, progesterone receptor (PR) positive and do not overexpress ERBB2. Luminal B

tumours are aggressive, high grade, ER and PR positive and do not overexpress *ERBB2*. Basal-like tumours do not overexpress *ERBB2* and are ER and PR negative. Some basal-like tumours however express one of these biomarkers (Sotiriou et al., 2003). ER negative tumours overexpressing the *ERBB2* gene are classified as the ERBB2 overexpressing group. Tumours with normal-like characteristics are grouped in the normal breast tissue-like subgroup. Patients with tumours classified as luminal A have the best prognosis, while patients with tumours classified as luminal B, basal-like and ERBB2 overexpression do poorly (Sorlie et al., 2001; Sorlie et al., 2003; Sotiriou et al., 2003; Hu et al., 2006). These subtypes represent various forms of the disease with different origins and it is thus important to understand their different molecular background in the hope of more accurate diagnosis and better treatment.

Prognostication and prediction of treatment effect of breast cancer

Therapeutic options vary depending on the molecular- and pathological composition of the tumour. Chemotherapy is often used in combination with radiation therapy and surgery (mastectomy). It targets all proliferating cells, but proliferation is one of the main properties of cancer cells. The use of chemotherapy in breast cancer treatment has been shown to be successful (EBCTCG, 2005) and mastectomy is a radical procedure where one or both breasts are removed surgically. Alternatively, some patients undergo lumpectomy where a small volume of the breast tissue is removed with some of the surrounding normal tissue for reassurance that no cancerous tissue is

left. No difference has been found between the two groups with respect to disease-free survival, distant-disease free survival or overall free survival (Fisher et al., 2002). Post-mastectomy radiotherapy has been shown to reduce relapse by at least two-thirds (Clarke et al., 2005; Rowell, 2009)

Targeted treatment has been developing over the past years, taking us one step closer to personalized medicine. Patients with Her2 overexpressing tumours are candidates for treatment with the anti-Her2 monoclonal antibody Trastuzumab (Hudis, 2007). Her2 overexpression/ERBB2 amplification can be detected by immunohistochemisty (IHC), fluorescence in situ hybridization (FISH), chromogenic in situ hybridization (CISH) and silverenhanced in situ hybridization (SISH) (Penault-Llorca et al., 2009). These methods are used to identify patients likely to benefit from treatment with the Her2 targeted monoclonal antibody Traztuzumab, which has been shown to have significant overall survival benefit (Smith et al., 2007). Trastuzumab (reviewed in (Goldenberg, 1999)) is a monoclonal antibody with two antigen-specific sites that bind the extracellular domain of the Her2 receptor, preventing the activation of its intracellular tyrosine kinase resulting in inhibition of cell growth (Carter et al., 1992). Hormonal therapy is available for patients with ER+ or PR+ tumours including Tamoxifen and aromatase inhibitors (reviewed in (Massarweh &Schiff, 2006)). Estrogen receptor positive tumour cells are activated by estrogen binding to the receptor thus promoting growth of the cells. Tamoxifen is a pro-drug that requires

metabolic activation to give rise to the active compounds that bind to the estrogen receptors inhibiting estrogen induced cell proliferation (reviewed in (Brauch &Jordan, 2009)). Until recently, patients with triple negative breast tumours (TNBT) had no targeted therapy available. TNBT and BRCA1 tumours share a similar genetic profile although TNBT patients do not necessarily carry mutations in the BRCA1 gene. However, alterations in the function of the BRCA1 gene product are believed to occur in TNBT patients (Miyoshi et al., 2008). Methylation at the promoter site leading to downregulation of expression is one such way (Catteau et al., 1999). The Parp and Brca proteins take part in repair of damaged DNA. In patients with *BRCA1/2* mutation, PARP inhibition prevents DNA repair, resulting in cell death (reviewed in (Comen &Robson)). Targeted breast cancer treatment is evolving and new therapeutic alternatives are still emerging. Continuation in the search for new therapeutic targets is of critical importance.

Genomic alterations in breast cancer

Breast cancer is a heterogeneous disease and is mainly the consequence of alterations of the genetic material, epigenetic factors and environmental factors. Alterations can give the cell additional properties and give rise to a more potent cell than its neighbours. However, some kind of equilibrium must be obtained since too many alterations can destroy the cell.

Hanahan and Weinberg (Hanahan &Weinberg, 2000) suggest a multistep process of tumourigenesis in human cells and that these

steps reflect genetic alterations that drive the progressive transformation of normal human cells into highly malignant derivatives. They suggest that for a cell to become malignant, six essential alterations need to occur (Figure 2); 1) Cells need to become self-sufficient in growth signals, normal cells require growth signals before they can move from a quiescent state into the proliferative state. Tumour cells can bypass that mechanism by activating the *H-RAS* oncogene, for example. 2) Cells need to become insensitive to growth inhibitors. Within a normal tissue, multiple antiproliferative signals operate to maintain cellular quiescence and tissue homeostasis, tumour cells can become insensitive to growth signals by losing the retinoblastoma suppressor, for example. 3) Tumour cells need to evade programmed cell death, for example by producing insulin-like growth factors (IGFs). The ability of tumour cells to expand in number is determined not only by the rate of cell proliferation but also by the rate of cell death. 4) Limitless replicative potential. One way for tumour cell to obtain that characteristic is to turn on their telomerase. Normal cells (except for stem cells) do not express telomerase and therefore their DNA shortens telomerically by each replication resulting in the cells´ death. 5) Limitless and sustained angiogenesis, for the transport of

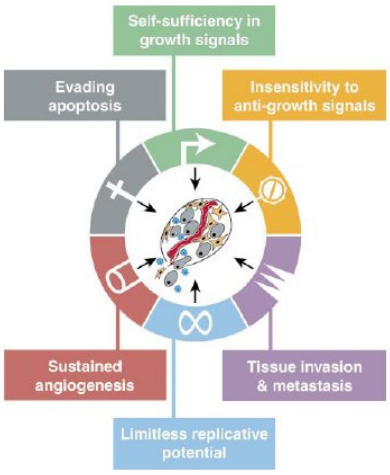


Figure 2 Acquired capabilities of cancer suggested by Hanahan and Weinberg (Hanahan &Weinberg, 2000).

oxygen and nutrition to the highly active tumour cells. That can be acquired by the production of vascular endothelial growth factors (VEGF's). 6) Tissue invasion and metastasis. Sooner or later in most malignant growth, tumour cells enter the bloodstream and travel to distant tissues where enough nutrition and space can be found, at least initially. For example, by inactivating E-cadherin the tumour cells free themselves from cell-cell adhesion (as reviewed in (Hanahan &Weinberg, 2000)) (Figure 2).

However, these six alterations are not sufficient to give a fully malignant phenotype as Hanahan and Weinberg point out in their article. The microenvironment has been shown to have an effect on tumourigenesis/malignant progression of the development of cells. The microenvironment of breast tumours is composed of many different cell types, for example adipocytes, fibroblasts, immune cells and endothelial cells. It has been reported that if non-tumourigenic mammary epithelial cells are transplanted in to mammary fat pads in irradiated mice, tumour formation is evident in 80% of hosts (Barcellos-Hoff & Ravani, 2000). These results indicate that radiation exposure can alter the microenvironment leading to neoplastic behaviour of the transplanted epithelial cells. On the other hand, normal fibroblasts have been shown to be able to convert malignant epithelial cells in the prostate to morphologically benign lesions (Hayashi &Cunha, 1991). Results from many studies indicate that the tumour microenvironment can modify the proliferation, survival, polarity, differentiation, invasive and metastatic capacities of cancer

cells (reviewed in (Hu & Polyak, 2008). Based on these results it is clear that tumour cells and cells in the environment collaborate and contribute to the malignancy.

Several mechanisms can cause these altered phenotypes of the malignant cells, e.g epigenetic changes, mutations, loss of heterozygosity, haploinsufficiency and an increase in DNA material. Epigenetic changes can alter the transcription of genes by either activation of oncogenes or inhibition of tumour suppressor genes. For example methylation of the cyclin-dependent kinase inhibitor *p16Ink4* gene, which is a tumour suppressor gene, leads to immortalization of breast and lung epithelial cells – one of the primary characteristics of malignant growth (Foster et al., 1998). Mutations can occur during transcription and give rise to abnormal mRNA and protein or even prevent transcription. Mutation in the tumour suppressor gene *PTEN* has been reported to have an important role in the carcinogenesis of breast cancer (Yang et al.). Loss of heterozygosity (LOH) is a mechanism where one allele is lost at a restricted area where the other allele is already inactivated, thus LOH can take part in inactivating tumour suppressor genes. LOH at the locus where the tumour suppressor genes *BRCA1* and *BRCA2* are located have been reported (Osorio et al., 2002). Haploinsufficiency is a state where one allele is inactivated and the other allele is not sufficient for mRNA and/or protein expression to maintain a normal phenotype, hence contributing to tumour formation. The tumour suppressor gene *PTEN* has been shown to be haploinsufficient in some tumour types (reviewed in

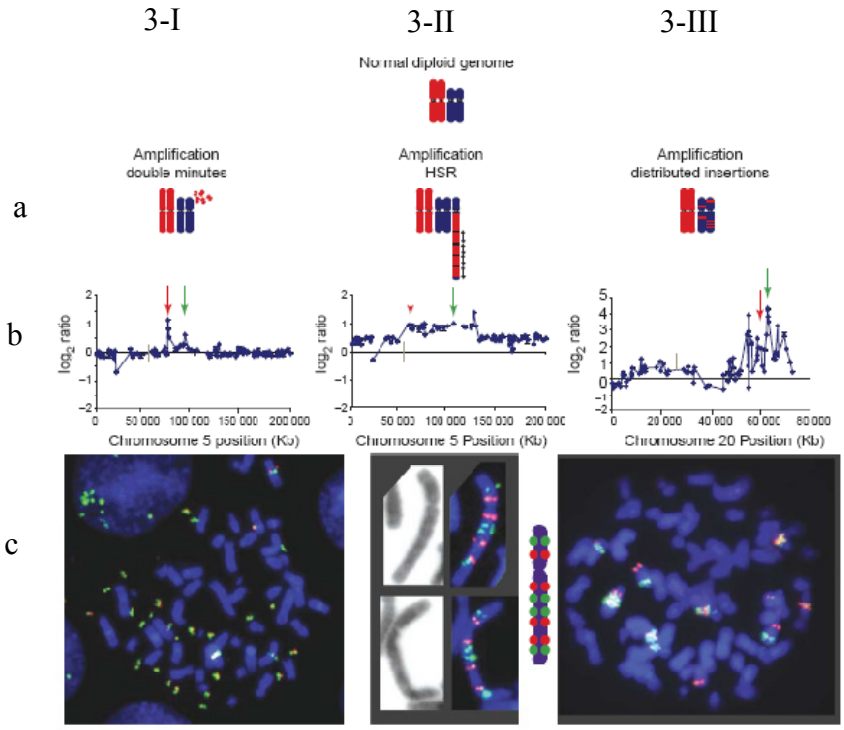
(Santarosa &Ashworth, 2004)). Increase (amplification and gain) of DNA material can also contribute to tumour formation in cases where oncogenes or genes taking part in pathways contributing to malignancy are amplified. Amplification of *CCND1* which encodes the cell cycle check point protein cyclin D is found in around 10% of breast tumours (Elsheikh et al., 2008). Overexpression of cyclin D has been shown to correlate with the amplification, resulting in limitless replicative potential (Reis-Filho et al., 2006a; Elsheikh et al., 2008).

Amplification and copy number increase in breast tumours

DNA amplification is a frequent event in tumour progression. The term DNA amplification refers to a copy number increase of a restricted region of a chromosome. Amplified regions can be from less than one megabase to tens of megabases in size. In human tumours, three different mechanisms are known to promote DNA amplification (Figure 3). In some cases the amplified region is entirely intrachromosomal, found repeatedly at the same chromosome as the original copy, called homogeneously staining region (HSR) (Figure 3-II) or scattered around the chromosomes (Figure 3-III). In other cases the amplified region is found extrachromosomally, called double minutes (Figure 3-I) (reviewed in (Albertson, 2006)).

Double minutes are extrachromosomal circular DNA fragments that replicate autonomously. Even though double minutes are acentric they segregate into daughter cells, by associating with mitotic chromosomes (Kanda et al., 2001). Some disagreement is on how double minutes are formed. Two mechanisms have been proposed.

The first one states that double minutes are formed by the "looping out" of chromosomal regions, generating circular DNA molecules and by unequal segregation leading to amplification (Toledo et al., 1993)



Donna G Albertson, Trends in Genetics 2006

Figure 3 Cytogenesis of DNA amplification. a) Schematic overview of the various forms of DNA amplification can be observed. As double minutes, homogeneously staining region (HSR) and distributed around the chromosome. b) Example of array CGH profile (log_2ratio as a function of chromosome position) for each type of amplification. c) FISH pictures of cells using BAC clones from the amplified region indicated by the red and green arrows. Left panel (3-I); double minutes are observed. Middle panel (3-II); HSR with inverted repeats. Right panel (3-III); amplified region scattered around the genome. As viewed in (Albertson, 2006).

The second proposes that double minutes are the consequences of double stranded breaks in cells lacking checkpoints (as reviewed in (Albertson, 2006)). Double minutes can reintegrate into chromosomal fragile sites giving rise to homogenously staining regions (HSR). It has been demonstrated that hypoxia is a potential fragile site inducer

that drives fusion of double minutes and their targeted reintegration site (Coquelle et al., 1998). Breakage at fragile sites can form inversions and HSR (Prentice et al., 2005). HSR is composed of a region found repeatedly within the same chromosome, a feature explained by the breakage-fusion-bridge cycle mechanism first described by Barbara McClintock in 1951 (McClintock, 1951).

Strong selective advantage is for tumour cells with amplification at regions encompassing genes that promote growth. Thus, genes located at amplified regions are potential oncogenes. This makes amplified regions excellent research tools for identifying new oncogenes. Also, amplification of a certain region has a clinical significance as well as patient prognosis, response to therapy and development of resistance (reviewed in (Albertson, 2006)). *ERBB2* is a good example of such gene.

ERBB2 is located at 17q12 and is found amplified in 30% of invasive breast tumours (Slamon et al., 1987). It encodes for epidermal growth factor receptor 2 (Her2) which takes part in signal transduction pathways leading to cell growth and differentiation (Olayioye, 2001). Amplification of the *ERBB2* gene leads to an increase in expression of the Her2 protein which can contribute to tumour formation (reviewed in (Olsen et al., 2004)).

Genomic amplification has been the subject of many breast cancer investigations. The most commonly detected amplified regions are 8p12-p11, 8q24, 11q13, 17q12 and 20q12 (Cuny et al., 2000; Letessier et al., 2006). The regions were first identified using methods

like Southern blotting (Cuny et al., 2000), metaphase chromosome CGH (Kallioniemi et al., 1994) and FISH (Al-Kuraya et al., 2004). More recent studies have confirmed those previous findings by applying more precise methods like array-CGH using BAC or oligo probes (Yang et al., 2004; Letessier et al., 2006; Kwek et al., 2009).

Genomic instability of chromosomal arm 8p in breast tumours

Genomic instability of the chromosomal arm 8p is frequent in malignancies derived from epithelial cells (carcinomas) especially breast carcinoma. Instabilities like loss of 8p21-ter, break in the 8p21-p12 region and amplification at 8p12-p11 are often observed. Also, there have been some indications of a fragile site lying within the gene *NRG1* (chr8: 32.6 – 32.7 Mb) located at 8p12 (Prentice et al., 2005).

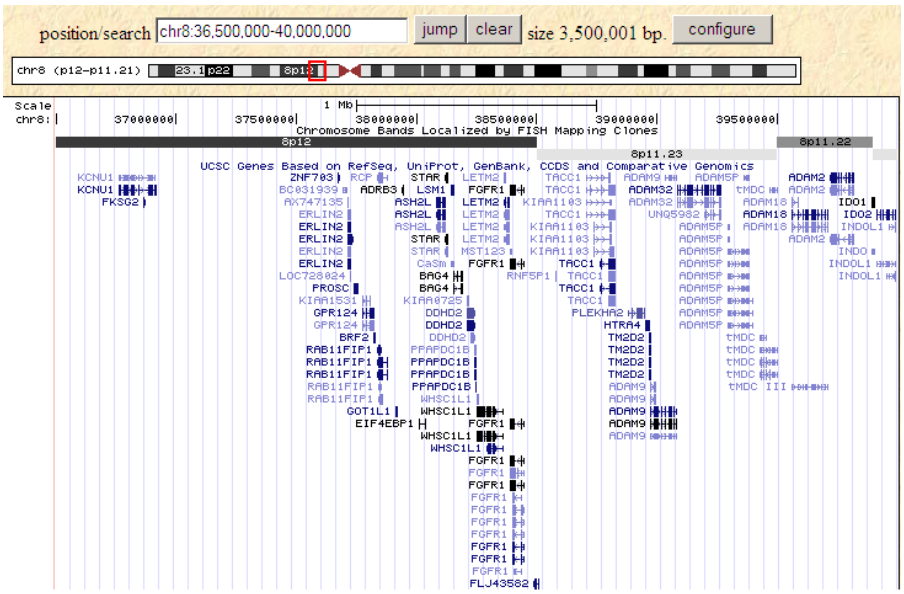


Figure 4 Overview of the gene arrangement of 3.5 Mb region at 8p12-p11 according to UCSC Genome Browser assembly (hg17).

Amplification and copy number increase of the 8p12-cen has been associated with poor prognosis (Prentice et al., 2005). Several research groups have studied the 8p12-p11 alteration in breast tumours and cell lines and most of them have located the main region of amplification spanning cytobands 8p12-8p11 (Figure 4). Some of the groups have identified a single amplified region (Garcia et al., 2005; Prentice et al., 2005) while others have mapped up to four separated amplified regions (Gelsi-Boyer et al., 2005; Kwek et al., 2009). The frequency of amplification varies from 7% to 25% in breast tumours, depending on cut-off levels and methods used for analysis (Adelaide et al., 1998; Ray et al., 2004; Garcia et al., 2005; Yang et al., 2006; Bernard-Pierrot et al., 2008).

Possible target genes of 8p12-p11 in breast tumours

Several studies have investigated the correlation between DNA copy number increase and mRNA expression for some of the genes located at 8p12-p11 in breast tumours and breast cancer cell lines. One study tested the correlation for 53 genes located within the 8p12-p11 region and observed positive correlation for 21 of them (Yang et al., 2006) while another study identified 24 of 64 tested genes having positive correlation (Gelsi-Boyer et al., 2005). Bernard-Pierrot et al. identified 11 genes of 28 genes tested at 8p12-p11 as constantly overexpressed when amplified (Bernard-Pierrot et al., 2008). Correlation between DNA level and mRNA amount is however only an indication of the gene taking part in tumourigenic pathways. It is important to test the genes further to narrow down the number of candidate oncogenes at

the 8p12-p11 amplified region. Many researchers have tested the tumourigenic characteristics of some of the genes by overexpressing them in non-tumourigenic cells or by knocking down their expression in tumour cell lines with amplification at 8p12-p11 (Yang et al., 2006; Streicher et al., 2007; Kwek et al., 2009; Zhang et al., 2009). Normally non-tumourigenic cells are not able to form colonies when grown in soft agar. By performing soft agar assay genes can be tested for their transformation properties by overexpressing them in non-tumourigenic cell lines. If the infected cells are able to form colony in soft agar then it is an indication of transformation potential of the gene also if the cells become independent of growth factors. By such an assay three genes located within 8p12-p11 (*LSM1*, *BAG4*, *RAB11FIP1)* have been identified to have transforming potential (Yang et al., 2006; Streicher et al., 2007; Zhang et al., 2009). In contrast to the results of these studies, another study identified 9 genes as not able to transform cells after their overexpression in non-tumourigenic cell line and therefore not induce colony formation in soft agar, amongst them were *LSM1*, *BAG4* and *RAB11FIP1* (Kwek et al., 2009). Another way of assessing the tumourigenic role of the genes on the 8p12-p11 amplified region is to knock down their expression in tumour cells harbouring the 8p12-p11 amplified region. If the infected cells slow down their proliferation or change their tumourigenic behaviour in some way it can be taken as an indication of the gene taking part in the tumourigenic pathways of the cells. Knocking down the expression of *PPAPDC1B*, *NSD3* and *LSM1* in tumour cells with amplification at 8p12-p11 reduced growth of the

cells dramatically, indicating their part in tumourigenic pathways (Streicher et al., 2007; Bernard-Pierrot et al., 2008). One of the genes located on the 8p12-p11 amplified region has been observed in an *in vivo* study (Bernard-Pierrot et al., 2008). Bernard-Pierrot et al. zenografted cells harbouring the 8p12-p11 amplified region in nude mice. Once tumours had established they were treated with PPAPDC1B-specific siRNA which bound to and promoted the brake down of the PPAPDC1B mRNA. Tumours treated with the PPAPDC1B-specific siRNA were 80% smaller than the control tumours, suggesting that amplification of *PPAPDC1B* contributes to oncogenesis and that the protein (encoded by *PPAPDC1B*) is a potential therapeutic target (Bernard-Pierrot et al., 2008). In the original studies on the 8p12-p11 amplified region *FGFR1* was considered the prime target gene (Adelaide et al., 1998; Ugolini et al., 1999). Several recent studies have ruled out *FGFR1* as the target gene of the amplified region based on that it is not always included in the amplified area and that inhibition of the fibroblast growth factor receptor fails to slow the growth of three breast cancer cell lines with 8p12-p11 amplification (Ray et al., 2004). However, *FGFR1* was identified as a potential target gene in a cell line with amplification at 8p12-p11 (MDA-MD-134) (Reis-Filho et al., 2006b). This may suggest that the oncogenic potential of *FGFR1* is restricted to certain breast cancer subtypes.

Taken together, a few genes have been identified more often than others as candidate oncogenes of the 8p12-p11 amplified region. They

are *PPAPDC1B*, *NSD3*, *LSM1*, *BAG4* and *RAB11FIP1* based on their transformation potential.

Other possible targets of the 8p12-p11 amplified region in breast tumours

Genes are not the only possible targets of the 8p12-p11 amplified region. Scattered around the genome are non-coding RNAs including microRNA (miR), small interfering RNAs (siRNA), small nuclear RNA (snRNA), small nucleolar RNAs (snoRNA) and even antisense of known genes (Mattick et al., 2006).

microRNAs are non-protein coding, consist of a single RNA strand about 21-23 nucleotides long and their function is to regulate gene expression. Alterations in microRNA expression have been linked to cancer pathophysiology (reviewed in (Shenouda &Alahari, 2009)). For example, increased expression of miR-10b has been linked to increased cell migration and invasion *in vitro* in breast tumour cells (Ma et al., 2007). microRNAs have been shown to map frequently at known fragile sites which can play a part in amplification and loss of restricted chromosomal regions (Calin et al., 2004). Comparison of normal breast tissue and breast cancer samples revealed that expression of microRNAs was aberrantly expressed in the tumour cells (Iorio et al., 2005) as well as correlated with *ERBB2* and *ER/PR* status (Mattick et al., 2006). Small interfering RNAs (siRNAs), produced in vitro for research and therapeutic purposes, function through conserved genetic supervision mechanism that takes part in post-transcriptional down-regulation of target genes. During the past

few years, siRNAs have been emerging as potential cancer therapy agents (reviewed in (Hadj-Slimane et al., 2007; Oh &Park, 2009)). Small nuclear RNAs (snRNA) function in the processing, translation and degradation of other RNA molecules (reviwed in (Jawdekar &Henry, 2008)), large group of snRNAs are known as small nucleolar RNAs (snoRNAs). The main function of snoRNAs is to guide site specific rRNA modification. snoRNA can be further classified into groups based on their function (as reviewed in (Dieci et al., 2009)). Some non-coding RNAs have been linked to malignant growth while others have not. Their function is complex and new information is emerging on their function and participation in cellular pathways.

Clinical relevance of the study

Improvement in genome wide analysis technology has made precise measurement of genome copy number aberrations possible. The increasing resolution of probes used for detection, can give more accurate information on DNA status of the whole genome. Amplified regions in the genome of cancer cells are likely to harbour genes participating in tumourigenesis. Therefore, they are of great interest when searching for new oncogenes. Also, amplification is thought to be the main mechanisms of resistance to therapy (reviwed in (Banerjee et al., 2002)).

The identification of the target genes of the 8p12-p11 amplified region could be of relevance as both a prognostic indicator and a drug target. As has been verified, the well established oncogene *ERBB2* is the

driver target gene of the 17q12 amplification and Trastuzumab, a monoclonal antibody drug which is used in cancer therapy.

Understanding and defining with more certainty what drives the 8p12-p11 amplification will increase knowledge and understanding of breast cancer development. It could increase the possibility to develop therapy against tumours with the amplification. Such therapy could improve the treatments already available, thus leading to better prognosis for a subgroup of breast cancer patients.

Aims of the study

The aims of the study were to map the minimal region of amplification at the 8p12-p11 chromosomal region in a large number of breast tumours and identify a potential target gene(s) within the region. In order to do so, DNA copy number data were used to map the copy number changes at the 8p12-p11 region in breast tumours and the minimal region of amplification was identified. For assessment of the potential target gene of the region the DNA copy number was correlated to the mRNA expression of the genes. Amplified genes with increased mRNA and protein expression are likely to take part in the tumourigenic pathways of a tumour cell. Therefore, protein expression of the genes with positive correlation between DNA copy number and mRNA expression was analysed in part of the tumours. Correlation of the protein expression to both DNA copy number and mRNA expression was analysed. The expression of the proteins was also analysed in cell lines with amplification at 8p12-p11 and compared to expression in non-malignant cells and cell lines without amplification. The genes that test positive in these studies will be identified as potential target genes of the 8p12-p11 amplified region.

Specific tasks:

1. Map the 8p12-p11 region in breast tumours and identify the minimal region of amplification
2. Analyse the correlation between DNA copy number and mRNA level of the genes within the minimal region of amplification

3. Quantify the protein expression of the genes with positive correlation between DNA copy number and mRNA expression and correlated to both DNA copy number and mRNA level
4. Analyse the protein expression of the genes showing positive correlation between DNA copy number and mRNA expression, in breast cancer cell lines and primary epithelial cells.

Materials and methods

Tumour tissue samples

Tumour tissue samples used in protein expression analysed in this study were collected at the Department of Pathology at Landspítali University Hospital. The tumour samples come from patients diagnosed with breast cancer in the years 1990-2001. The research is part of a long term project approved by The Icelandic Data Protection Committee and the Ethical Committee.

Cell lines and protein from primary breast epithelial cells

Cell lines were selected based on their DNA status according to published data (for reference see Table 7). Cell lines with amplification/gain and loss at chromosomal region 8p12-p11 were chosen along with one non-tumourigenic cell line with normal DNA status at 8p12-p11. A total of 7 cell lines were selected for the study (SUM52, SUM44, CAMA-1, HCC1500, MCF-7, MCF10A and MDA-MB-231).

SUM52 and SUM44 were purchased from Asterand. CAMA-1 and HCC1500 were purchased from the American Type Culture Collection (ATCC). MCF-7, MCF10A and MDA-MB-231 were a gift from Dr. Þórarinn Guðjónsson.

Proteins from two types of primary breast epithelial cells were included in the study. Proteins from myo-epithelial and luminal-epithelial cells were a gift from Dr. Þórarinn Guðjónsson. The cells

were isolated from reduction mammoplasties and processed as described (Pechoux et al., 1999; Gudjonsson et al., 2002).

Cell line culture conditions

SUM52 and SUM44 breast cancer cell lines were cultured in DMEM/F-12 based medium (Gibco-Invitrogen) with 5% FBS (Fetal bovine serum) and supplemented with 50 IU/ml penicillin, 50 µg/ml streptomycin, 5 µg/ml insulin and 1 µg/ml hydrocortisone. MCF10A was cultured in a chemically defined medium H14 consisting of DMEM/F12 medium supplemented with 50 IU/ml penicillin, 50 µg/ml streptomycin, 10 ng/ml EGF, 0.1 nM estradiol, 0.5 µg/ml hydrocortisone, 250 ng/ml insulin, 2.6 ng/ml Na_2SeO_3, 5 µg/ml prolactin and 10 µg/ml transferrin. The HCC1500 cell line was cultured in RPMI-1640 medium (Gibco-Invitrogen) with 10% FBS and supplemented with 50 IU/ml penicillin, 50 µg/ml streptomycin and 10 mM HEPES. The MDA-MB-231 cell line was cultured in RPMI-1640 medium (Gibco-Invitrogen) with 10% FBS supplemented with 50 IU/ml penicillin and 50 µg/ml streptomycin. The CAMA-1 cell line was cultured in DMEM/F-12 with 10% FBS and supplemented with 50 IU/ml penicillin and 50 µg/ml streptomycin and 0.1 mM NEAA (Non essentials amino acids). The MCF7 cell line was cultured in DMEM/F-12 with 10% FBS and supplemented with 50 IU/ml penicillin and 50 µg/ml streptomycin. All cells were cultured at 37°C in a humidified incubator at 5% CO_2.

Analysis of data from the array-CGH

The array CGH data used in this study were accessible though BASE (http://base.onk.lu.se/onk/), which is a databank run by the University Hospital of Lund in Sweden. In the years 2004 - 2006 Haukur Gunnarsson, Rósa Björk Barkardóttir and Aðalgeir Arason, all members of the Department of Pathology performed array-CGH on 217 Icelandic breast tumours as described Jonssoon et al. (Jonsson et al., 2007). Briefly, tumour DNA and normal male reference DNA were labelled with Cy5 and Cy3, respectively and amplified by polymerization reaction. The differentially labelled tumour and normal DNA were then mixed along with COT-1 DNA to reduce unspecific binding. After the samples had been dried and re-suspended in hybridization buffer the mixture was hybridized on microarray slides. The microarray slides were high-resolution tiling arrays containing over 32000 BAC clones with the average resolution of 80 kb over the genome. The microarrays were then scanned to detect fluorescent signals from the dyes. Poor quality data spots were flagged in GenePix software before the data was imported to BioArray Software Environment (BASE, http://base.onk.lu.se/onk/), a web-based software package for storing, searching and analysing locally generated microarray data and information surrounding microarray production (Saal et al., 2002).

In this study the aCGH data obtained from these 217 breast tumours were further analysed according to (Staaf et al., 2007). The data was corrected for background signal by subtracting the median background

from the median-foreground signal intensities for each channel and $\log_2$ ratios calculated for each spot. Data of poor quality that had been flagged in GenePix were filtered out along with spots with signal to noise ratio (SNR) lower then 5. Paraffin embedded tumour samples had been hybridized twice, once with each dye. Thus, results from the two dyes were merged by using the average intensity of the spots. Spots of intensity greater than 2 standard deviations from the mean were eliminated. The data was normalised using Normalisation Poplowess (Staaf et al., 2007), which is an algorithm designed to normalise data displaying numerous copy number variations. Clones bound to chromosomes X and Y were excluded.

Data from the GEx study

The GEx data used in this study were accessible from the same databank as the aCGH data (BASE, http://base.onk.lu.se/onk/). Gene expression of 161 primary tumour samples from Icelandic patients were analysed by array Gene Expression (GEx) and the data normalised. This was performed by Aðalgeir Arason and Rósa Björk Barkardóttir as described by Staff et al. (Staaf et al., 2010). The GEx arrays included ~35000 probes each. The probes on the arrays did not in some cases detect all the known transcripts of the genes.

Mapping of chromosomal regions

Mapping of chromosomal arm 8p

For mapping of chromosomal arm 8p Mev (Multiple array viewer) v4.0 software was used. Mev is available for download online at http://www.tm4.org/mev/. Data were extracted from BASE and

imported to Mev where the data were cleaned before they were plotted, only samples with results for over 80% of clones were plotted.

Mapping of chromosomal region 8p12-p11

For mapping of the chromosomal region 8p12-p11, software specially designed for this study was used. The software was designed by Elena Losievskaja and called Map. Data was extracted from BASE and imported in to Map for plotting. The picture obtained from Map was used for visual aid only. The exact minimal region of amplification was determined by examine the data in BASE for exact basepair location of the cores of the amplified regions.

Protein extraction

Cell lines

Proteins were extracted from the cells straight from the culture flasks. Cells were washed twice with cold 0.1M PBS pH 7.4 (8 mg/ml NaCl, 0.2 mg/ml KCl, 7.62 mg/ml Na_2HPO_4 and 0.77 mg/ml KH_2PO_4). The cells were covered with RIPA lysis buffer (150 mM NaCl, 50 mM Tris pH 7, 1.0% NP40, 1.0% Na-deoxycholae and 0.1% SDS), containing 1x protease inhibitor (PI) and the flask was kept on ice for 10 minutes. Then the cells were scraped off the (inner surface) bottom of the flask with a cell scraper and transferred to a test tube. The whole-cell lysate was squeezed though a needle several times to break down DNA of high molecular weight and the cell lysat kept on ice for another 10 minutes. The samples were centrifuged for 20 minutes at 4°C and the supernatant was used for the analysis.

Breast tumours

The tumours, which are stored in aluminium bags at -80°C, were submerged in liquid nitrogen to reduce proteinase activation. A piece of the tumour was cut using a scalpel on a plate placed on dry ice. The piece was placed in an aluminium bag and submerged in liquid nitrogen (LN_2) at around -196°C then placed in a mortar where it was crushed. The crushed tumour was then placed in a beaker pre cooled in LN_2 and placed in the Micro Dismembrator S (Sartorius) for 30 seconds at 2600 rpm. Then, the beaker was cooled in the LN_2 and put in the dismembrator for another 30 seconds. Subsequently the micro-dissected tumour was dissolved in 300-400 µl RIPA buffer (containing 1x protease inhibitor cocktail). The whole-cell lysate was moved to a test tube and squeezed though a needle to break down DNA of high molecular weight and then centrifuged for 20 minutes at 4°C. The concentration of the proteins, found in the supernatant, were then determined.

Determination of protein concentration

The protein concentration of the lysates was determined by using Pierce BCA Protein Assay Kit (Thermo Scientific). The BCA assay (bicinchoninic acid assay) is based on the reduction of Cu^{2+} to Cu^{+} in the presence of protein under alkaline conditions. The amount of reduction is proportional to the amount of proteins present. Then Cu^{+} ion binds 2 BCA molecules in an alkaline environment which is a purple coloured reaction (equations 1). The Cu^{+}BCA complex exhibits strong absorbance at 562 nm which was detected in Nano Drop 1000

spectrophotometer. The working reagent solutions were mixed according to the manufacture´s protocol and 200 µl added to 10 µl of the tumour samples. The mixture was heated at 37°C in a waterbath for 30 minutes and cooled to room temperature. The protein concentration was then measured.

Equation 1

$$1\text{x protein} + 1\text{x } Cu^{2+} \rightarrow 1x\ Cu^{+}$$

$$1\text{x } CU^{+} + 2\text{x BCA} \rightarrow 2BCA - Cu^{+}\ complex$$

Western blotting

Sample preparation

At the beginning of the study 20 µg of proteins were used in each well on the SDS-PAGE gel but after a few Western blots the quantity was lowered to 10 µg and then to 5 µg per well. The proteins were mixed with 1x Protein samples buffer (PSB) (60mM Tris, pH 6.8, 2% SDS, 10% glycerol, 0.008% bromphenol blue and 5% beta-mercaptoethanol). The SDS binds to the proteins and gives them negative charge, beta-mercaptoethanol removes disulfide bridges from the proteins, glycerol gives the proteins more weight so they will not float out of the wells on the SDS-PAGE gels and bromophenol blue gives the proteins blue color for visualisation of the proteins. The samples were heated at 95°C for 5-10 minutes before loading in gels. Except, samples loaded on gels for detection of the membrane bound protein Rab11fip1 were heated at 60°C for 10 minutes. Membrane

bound proteins can aggregate when heated at 95°C making running through the gel difficult.

SDS-PAGE gels

SDS-PAGE (sodium dodecyl sulphate - polyacrylamide gel elecrophoesis) of different concentrations (7.5% - 15% acrylamide) gels were cast depending on the predicted size of the proteins of interest. Recipe for 10 ml of a 10% acrylamide gel: 4 ml H_2O, 3.3 ml 30% acrylamide mix (29.2% acrylamide and 0.8% N,N'methylene-bis-acrylamide), 2.5 ml 1.5 M Tris pH 8.8, 0.1 ml 10% SDS, 0.1 ml 10% APS (ammonium persulfate) and 4 µl TEMED. For increased or decreased acrylamide concentration the volume of acrylamide and H_2O were changed accordingly.

SDS is anionic detergent and imparts negative charge to all the proteins in the sample rendering every protein the same charge-to-mass ratio. Therefore and because of obstacle in the gel, the mobility becomes a function of molecular weight. Thus, the velocity of the charged protein moving in an electrical field is directly proportional to the size of the molecule. Hence, increased acrylamide concentration increases the difference in mobility between proteins of different molecular weight.

The gels are composed of a stacking gel (5% acrylamide) and a running gel (7.5-15% acrylamide). The proteins were loaded in the wells of the stacking gel. The purpose of the stacking gel is to line up the proteins for them to enter the running gel at the same time.

Running of gels

The gels were placed in a vertical electrophoresis cell (Mini PROTEAN Tetra Cell, BioRad) and filled up with running buffer (3.03 mg/ml Tris, 14.4 mg/ml Glycine, 10 mg/ml SDS). Volume of 6 - 15 µl of protein samples (5-20 µg) and PSB was loaded in the wells using microliter syringe (Hamilton). The vertical electrophoresis cell was connected to a power supply (BioRad Power Pac 3000 or GibcoBRL PS 304) and the proteins run through the stacker gel at 100 V and through the running gel at 150 V.

Each of the proteins was detected in total of 52 samples (39 breast tumour samples, 11 cell lines and 2 types of primary epithelial cells). Therefore, 4 gels with 15 wells were loaded at the same time. For determination of the size of the proteins 0.5 - 1 µg of a Page Ruler prestained protein ladder (Fermentas) was loaded in one well of each gel. The cell line SUM52 was also loaded on each gel for comparison of protein amount between gels (see "Analysis of protein amount data" in Materials and methods).

Transfer of proteins to PVDF membranes

Proteins were electrophoretically transferred in TRANS BLOT, Semi dry transfer cell (BioRad) to an Immobilon-P PVDF (Polyvinylidene Difluoride) membrane with a pore size 0.45 µm (Sigma-Aldrich). The PVDF membrane was pre-wetted in methanol and washed with H_2O and transfer buffer (30 mg/ml Tris and 144 mg/ml Glycine) before the transfer. The membrane was placed between six sheets (three on each side) of Whatman (3 mm) paper, wetted in transfer buffer and placed

in the TRANS BLOT. The TRANS BLOT was connected to power supply and transferred for 90 minutes at 15 V. The transfer of the proteins was verified by soaking the membrane in 0.1% Ponceau S.

Probing with antibodies

Commercially available antibodies were purchased for detection of the proteins of interest. Many of the proteins have more than one known isoform but the available antibodies only recognised one or two of them.

To prevent non-specific background binding of the primary and secondary antibodies the membrane was blocked with 5% non-fat milk (mixed in 1x TBST (1.2 mg/ml Tris, 8.77 mg/ml Glycine and 0.05% Tween)) by incubating for 60 minutes on an orbital shaker at 10-15°C.

The primary antibodies (Table 1) specific to the target protein were diluted in incubation buffer containing (5% non-fat milk in 1x TBST) and the membranes incubated over night at 10-15°C on an orbital shaker. Membranes were rinsed 4x for total of 20 minutes with 1x TBST before hybridization with a secondary antibody.

Horseradish peroxidase (HRP) labelled secondary antibodies were used to bind to the primary antibody. The secondary antibodies (Table 2) was also diluted in incubation buffer containing (5% non-fat milk in 1x TBST) and the membrane incubated for one hour at 10-15°C on an orbital shaker. The membranes were rinsed 4x for total of 20 minutes before detection of the protein bands.

Table 1. Information about primary antibodies used in the study.

Antibody	Used at concentration	Clonality	Raised in	Purchased from
anti-Znf703	0.25 µg/ml	Polyclonal	Goat	Abcam
anti-Spfh2	1.00 µg/ml	Polyclonal	Goat	Abcam
anti-Prosc	0.10 µg/ml	Polyclonal	Rabbit	SDI
anti-Brf2	0.50 µg/ml	Polyclonal	Goat	Abcam
anti-Rab11fip1	0.25 µg/ml	Polyclonal	Goat	Abcam
anti-Ash2l	2.00 µg/ml	Monoclonal	Mouse	Abcam
andi-Lsm1	0.20 µg/ml	Polyclonal	Rabbit	SDI
anti-Bag4	0.20 µg/ml	Polyclonal	Rabbit	SCBI
anti-Star	0.25 µg/ml	Polyclonal	Rabbit	Abcam
anti-Ddhd2	0.15 µg/ml	Polyclonal	Rabbit	SDI
anti-Nsd3 (72.6 kDa)	0.25 µg/ml	Polyclonal	Rabbit	Abcam
anti-Nsd3 (165 kDa)	0.30 µg/ml	Polyclonal	Rabbit	Novus Biol.
anti-Fgfr1	1.00 µg/ml	Polyclonal	Goat	Abcam
anti-Actin (C4 clone)	0.034 µg/ml	Monoclonal	Mouse	ICN
anti-Gapdh	0.057 µg/ml	Monoclonal	Mouse	RDI

SDI; Strategic Diagnostics Inc, **SCBI**; Santa Cruz Biotechnology INC, **ICN**; ICN Biomedicals, **RDI**; RDI Division of Fitzgerald Industries Intl, **Novus Biol**; Novus Biologicals.

Table 2. Information about secondary antibodies used in the study.

Antibody	Used at concentration	Raised in	Purchased from
anti-mouse	0.05 µg/ml	Goat	DakoCytomation
anti-goat	0.075 µg/ml	Rabbit	Invitrogen
anti-rabbit	0.075 µg/ml	Goat	Invitrogen

The antibodies and proteins were stripped of the membrane before probed with loading control using Stripping buffer pH 2.5 (15 mg/ml Glycine and 0.1% Tween) for 20 minutes on orbital shaker and washed 4x for total of 20 minutes with TBST before new antibodies were hybridised.

Detection of protein bands

Proteins were detected by incubating the membranes with ECL Plus (GE Healthcare; 0.1 ml/cm^2) solution in the ratio 40:1 of the BCA protein assay solutions, respectively, according to the manufacturer for 2 minutes at RT. The excess ECL was drained off and the membrane placed in a plastic wrap with the protein side up, closed up and placed in an x-ray film cassette. In a darkroom, the membranes were exposed to a photographic film (GE Healthcare) for various length of time and developed in the film processor (X-OMAT and DURR MED 260).

The ECL Plus solution contains amongst other ingredients Lumigen PS-3 which reacts with the horseradish peroxidise on the secondary antibody to yield CO_2 and light (chemiluminescence). Multiple exposure times were performed to ensure results were obtained within the linear range of the film.

Quantification of protein bands

The films were scanned in CanoScan LiDE 90 (Canon) in greyscale, resolution 600 dpi and in .tiff format. The quantification was performed in Photoshop CS2 (Adobe). The protein bands that developed on the films are represented as dark bands, thus the higher the protein amount, the darker the bands. Photoshop reports dark points near null and light points have values that max out at 255. Therefore, the image was inverted so that darker bands (lighter after inversion) will get high numerical values when measured compared to weaker bands (darker after inversion) that get lower numerical values.

Using the lasso tool in Photoshop a line was drawn manually around the edges of the protein bands. The mean value and the pixel value of the selected area were measured. The mean value, which gives information on the mean grey value (from 0 – 255) of the selected area, and the pixel value, which gives information of how many pixels are in the selected area, were measured. The mean value of a same size area was measured for the background above or under each protein band. The data was collected in Excel sheet for analysis.

Since the line was drown manually around the protein bands, over- or under assessment of the protein bands was at risk. For assessing the consistency of measurement, three persons measured the bands in all the samples for one protein and obtained comparable results.

Analysis of protein quantification data

The mean value of each protein band, mean value for the background of each band and pixel value for both (same pixel value for both) was used for the assessment of the intensity of each band to obtain the absolute intensity of the protein of interest. The mean value for the background was subtracted from the mean value of the measured protein band and then multiplied by the pixel value. To correct for human and technical factors, e.g protein loading, transfer efficiency and different time of exposure, housekeeping proteins (Gapdh or β-actin) were analysed in the same way as the proteins of interest. The absolute intensity of the protein of interest was divided by the absolute intensity of the housekeeping protein and the corrected absolute intensity obtained for the protein of interest (equation 2). Since the

samples, cell lines and primary epithelial cells were loaded on a total of 4 gels the results was normalized for comparison between gels. Therefore, the cell line SUM52 was loaded on each gel. The corrected absolute value for all the samples on each gel was divided with the corrected absolute value for SUM52 on the same gel to come up with the relative intensity (Equation 3). Thus, the results for the same protein on different gels were comparable. The relative intensity is a unitless value which was $\log_2$ transformed for statistical calculations.

Equation 2

$$(\text{mean} * \text{pixel})(\text{POI}) - (\text{mean} * \text{pixel})(\text{Bg}) = absolute\ intensity\ (POI)$$

$$(\text{mean} * \text{pixel})(\text{HP}) - (\text{mean} * \text{pixel})(\text{Bg}) = absolute\ intensity\ (HP)$$

$$\frac{\text{absolute intensity (POI)}}{\text{absolute intensity (HP)}} = corrected\ absolute\ intensity\ (POI)$$

Equation 3

$$\frac{\text{corrected absolute intensity (POI) SAMPLE}}{\text{corrected absolute intensity (POI) SUM52}} = relative\ absolute\ intensity\ (POI)\ SAMPLE$$

POI= Protein of interest, HP= Housekeeping protein, Bg= Background

Statistical analysis

All statistical analysis was performed in Microsoft Office Excel 2007. Statistical significance (p-value) was tested in open access at http://www.danielsoper.com/statcalc/. Correlation calculation between DNA level, mRNA amount and protein amount was done using Pearson product moment correlation coefficient (r) function in Excel.

Log_2 transformed values were used for the calculations and was estimated statistically significant if the p-value was lower than 0.05.

RESULTS

At the Department of Pathology two data banks are available based on results from studies using microarray technology to detect DNA and mRNA levels in breast tumours. Genetic copy number changes were analysed in the whole genome of 217 breast tumours from Icelandic women using microarray comparative genomic hybridization (aCGH). Gene expression analysis was performed in 161 of the same tumour samples on oligo based microarrays (GEx). The aCGH study was done by Haukur Gunnarsson, Aðalgeir Arason and Rósa Björk Barkardóttir and the GEx study was performed by Aðalgeir Arason and Rósa Björk Barkardóttir over the years 2004-2006. Both the aCGH and the GEx studies were done in collaboration with Professor Åke Borg and his team at the University Hospital in Lund, Sweden. The arrays and the techniques used in these studies were described in previously published papers (Jonsson et al., 2007; Staaf et al.). In this study the aCGH data bank was used to map copy number alterations in 217 Icelandic breast tumours. Increases in copy number were defined as gain when the ratio of the tumour DNA to the normal male reference DNA was ≥ 1.2 ($\log_2 \text{ratio} \geq 0.26$). Amplification and high amplification were defined as ratio ≥ 1.6 ($\log_2 \text{ratio} \geq 0.68$) and ratio ≥ 3 ($\log_2 \text{ratio} \geq 1.58$), respectively. Decrease or loss of copy number was defined as ratio < 0.8 ($\log_2 \text{ratio} < -0.32$). It should be noted, that a ratio that equals to or is higher than 1.6 is the equivalent of 2 or more extra copies of the amplified DNA region, assuming the tumour samples are composed of 80% tumour cells and 20% non-tumour cells (immune cells, fibroblasts, endothelial cells etc.).

Likewise, ratio that equals 3 or more is equivalent to 3.8 copies or more of the amplified DNA region.

In this study, genes located at a known amplified region at 8p12-p11 were analysed in breast tumours, breast cancer cell lines and primary epithelial cells. The genes must meet four criteria to be considered candidate target genes of the amplification. In breast tumours, they must be overexpressed at the mRNA level when amplified at the DNA level, positive correlation must be observed between protein expression and mRNA expression and also DNA copy number, and at least 3 of the 4 cell lines with amplification at 8p12-p11 must express higher amount of the protein encoded by the gene of interest than cell lines with normal DNA or loss at the 8p12-p11 region.

The aCGH data was used to define the most frequently amplified (ratio ≥ 1.6) regions in the 217 Icelandic breast tumours. The most frequently amplified regions were 17q12 (15%), 8p12 (14%) and 11q13 (14%). The well defined oncogene *ERBB2* is located at 17q12 and is found amplified in a large proportion of breast tumours (Slamon et al., 1987). *CCND1* which encodes the cell cycle check point protein cyclin D1 is located at 11q13 and is known to be amplified in many breast tumours (Reis-Filho et al., 2006a). The amplification at 8p12-p11 is a well defined alteration in breast tumours but the driver target gene has not been identified yet. Therefore, the 8p12-p11 amplified region was selected for further studies.

Frequency of copy number changes at chromosomal arm 8p

aCGH data from the 217 tumour samples was used to map copy number changes at chromosomal arm 8p. A total of 449 BACs provide near tiling path coverage of a 44 Mb region of 8p, with an average of 100 kb resolution.

To get an overview of the copy number imbalance of the chromosomal arm, samples were classified into groups depending on their level of copy number at 8p according to the aCGH data. For observation of copy number increase the samples were classified by their highest ratio at 8p. Regions are considered highly amplified, amplified or having gain when the ratio was equal to or greater than 3, 1.6 and 1.2, respectively (Table 3). For observation of copy number decrease the samples were classified by their lowest ratio at 8p. Regions were considered having loss and extreme loss when ratio was lower than 0.8 and 0.5, respectively (Table 4).

Observed frequency of copy number changes was high within chromosomal arm 8p. Copy number imbalance (decrease or increase) was detected in 121 tumours (56%). Copy number increase (ratio $\geq$ 1.2) was observed in 86 samples that represent total of 40% of the tumours (86/217). Of the 86 tumours with copy number increase, 46 sample (21% of total) had gain ($1.2 \leq$ ratio < 1.6), 21 samples (10% of total) had amplification ($1.6 \leq$ ratio < 3) and 19 samples (9% of total) had high amplification ($3 \leq$ ratio) (Table 3). Of the 217 tumour samples 83 (38% of total) showed decrease in copy number at 8p. Samples with loss and extreme loss were 77 and 6, respectively.

Samples showing both decrease and increase at 8p were 48 (22% of total).

Table 3. Frequency of copy number increase at 8p in breast tumours

Group	Definition	Number of samples	Percentage
High amplification	$3 \leq$ ratio	19	9%
Amplification	$1.6 \leq$ ratio < 3	21	10%
Gain	$1.2 \leq$ ratio < 1.6	46	21%
		86	40%

The samples were classified by their highest ratio at 8p. aCGH data from 217 samples was used for the frequency calculations. Total of 86 samples (40%) have copy number increase at 8p.

Table 4. Frequency of copy number decrease at 8p in breast tumours

Group	Definition	Number of samples	Percentage
Loss	$0.5 \leq$ ratio < 0.8	77	35%
Extreme loss	ratio < 0.5	6	3%
		83	38%

The samples were classified by their lowest ratio at 8p. aCGH data from 217 samples was used for the frequency calculations. Total of 83 samples (38%) have decrease in copy number at 8p.

Mapping of copy number changes at 8p

Mapping of copy number changes in the tumour cells revealed frequent loss of distal 8p and gain and/or amplification at the region next to the centromere (figure 5). The region with the highest frequency of amplification was found at 8p12-p11.

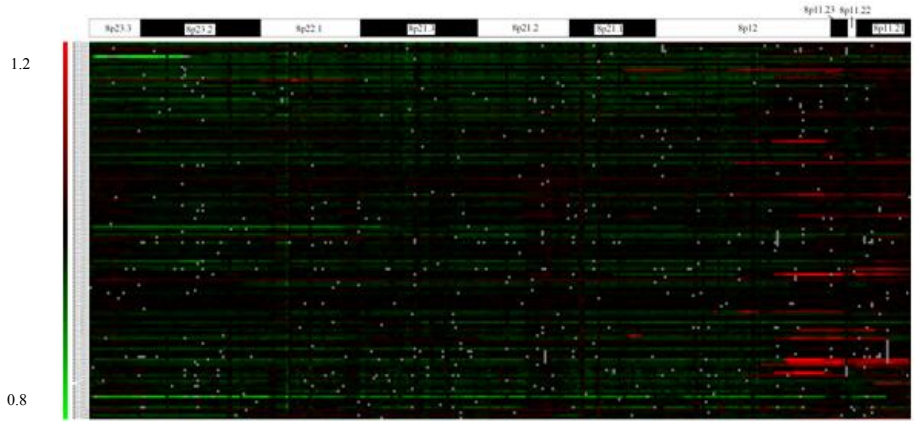


Figure 5. Frequent amplification centromeric at 8p in breast tumours. Copy number changes at 8p detected by aCGH. Red; increase in copy number ($1.2 \leq$ ratio), green; decrease in copy number (ratio < 0.8) and black; normal DNA level ($0.8 \leq$ ratio < 1.2). Grey reflects missing data. Tumour samples are aligned in rows along the 8p arm. On top; a schematic of 8p is depicted with the cytobands labelled telomeric (left) to centromeric (right).

Mapping of DNA copy number changes at 8p12-p11

In tumour formation strong selection is for cells with increased level of genes that promote tumour growth. Therefore, genes located at amplified regions are possible oncogenes. In order to define the minimal region of amplification at 8p12-p11, the DNA profiles of the 40 samples that showed amplification and high amplification (ratio ≥ 1.6) at 8p were aligned.

The minimal region most often amplified was identified at 37.3 to 38.9 Mb defined telomeric by S1 and centromeric by S24 and S26 (Figure 6). The region can possibly be divided in two smaller amplifications. Despite that, the minimal region of amplification was considered as one continuous region and all the genes at the region included for further studies. Thus, the minimal region of amplification was identified as 1.6 Mb in size and

encompassed 20 known protein coding genes; *LOC157860, ZNF703, SPFH2/ERLIN2, PROSC, GPR124, BRF2, RAB11FIP1, GOT1L1, ADRB3, EIF4EBP1, ASH2L, STAR, LSM1, BAG4, DDHD2, PPAPDC1B, NSD3/WHSC1L1, LETM2, FGFR1* and *TACC1* according to the UCSC Human Genome Browser Build 35 (hg17) and Ensembl (release 56).

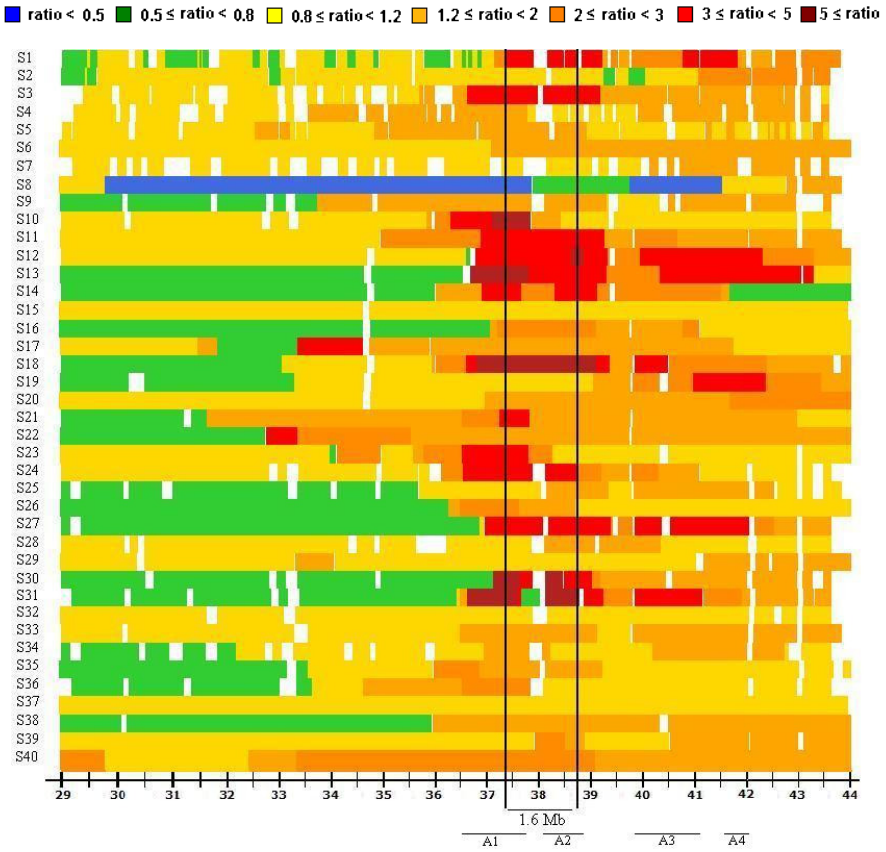


Figure 6. The minimal region of amplification in the breast tumours was 1.6 Mb. Mapping of the chromosomal region 8p12-p11 in the breast tumours with amplification (ratio ≥ 1.6) at 8p. The minimal region was defined as the region which most samples had their highest point of amplification. Most samples had their highest point of amplification within the region from 37.3 – 38.9 Mb. Under the Mb scale are marked the four amplified regions identified by Gelsi-Boyer et al. indicated as A1-A4 (Gelsi-Boyer et al., 2005). See definitions for color coding above figure, white; data not available.

Table 5. Copy number increase at the 1.6 Mb minimal region of amplification

Group	Definition	number of samples	Percentage
High amplification	ratio ≥ 3	17	8%
Amplification	1.6 ≤ ratio < 3	21	10%
Gain	1.2 ≤ ratio < 1.6	32	15%
		70	32%

The samples were classified after the highest ratio observed at the 1.6 Mb region. Therefore only copy number increase is displayed in the table. Total of 217 samples were examined for this frequency calculations and thereof 70 samples (32%) displayed copy number increase.

Frequency of copy number increase at the 1.6 Mb minimal region of amplification was observed according to the aCGH data for the 217 breast tumours. A total of 70 samples (32%) had copy number increase (ratio ≥ 1.2) at the 8p12-p11 amplified region. Thereof 32 samples (15% of total) had gain (1.2 ≤ ratio < 1.6), 21 samples (10% of total) had amplification (1.6 ≤ ratio < 3) and 17 samples (8% of total) had high amplification (Table 5).

Correlation of DNA copy number and mRNA expression of genes located within the minimal region of amplification

Genes at amplified regions that exhibit increased mRNA and protein expression are likely candidate oncogenes. To analyze whether the mRNA expression was enhanced from genes residing at the 8p12-p11 amplified region, expression data were extracted from the gene expression data bank (GEx data bank). Data for mRNA expression were available for 17 of the 20 genes located within the minimal region. The 17 genes are: *ZNF703, SPFH2, PROSC, GPR124, BRF2, RAB11FIP1, ADRB3, EIF4EBP1, ASH2L, LSM1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, LETM2, FGFR1* and *TACC1*. The results for *STAR* were not

of sufficient quality to use in the analysis and *LOC157860* and *GOT1L1* were not included on the gene expression arrays.

In this study, positive correlation between DNA copy number and mRNA expression is one of the criteria set for a gene to be considered a candidate target gene of the amplified region. Data of sufficient quality for both DNA and mRNA analysis were available for 152 samples (although the number of samples may differ between genes) and for 17 of the 20 genes located at the minimal region of amplification. Correlation was calculated using Pearson product-moment correlation coefficient (r) and results were considered statistically significant if the p-value was lower than 0.05.

Thirteen of the genes encode for more than one transcript (all but *BAG4*, *LSM1, ADRB3 and EIF4EBP1*). For four of the 17 genes, the GEx probe identified more than one transcript of the gene (*GPR124, ASH2L, NSD3* and *TACC1*). It should be noted that the probe that identified expression of *NSD3* binds one transcript of *NSD3* and one transcript encoded by another gene in the same gene-family, *NSD1*. In that case the results of mRNA level represent the level of both of the transcripts the probe binds. This multiple identification of probes was not relevant when comparing DNA level and mRNA level but has effect on the comparison of mRNA level and protein amount. For one gene, three probes bind to the same transcript (*RAB11FIP1*). In that case, the average value was used. For one gene (*FGFR1)* both of the above apply, that is three probes bind to 4-5 transcripts of the gene. In that case, the average value of the three probes was used for DNA and

mRNA correlation calculations. The correlation between mRNA and protein level could not be performed because none of the probes identified the transcript that encodes for the isoform recognised by the antibody.

Table 6. Correlation of DNA copy number and mRNA expression of the genes located at the minimal region of amplification.

	r	n	p-value
ZNF703	0.53	152	< 0.001
SPFH2	0.77	152	< 0.001
PROSC	0.84	151	< 0.001
GPR124	-0.05	152	0.541
BRF2	0.67	51	< 0.001
RAB11FIP1	0.62	151	< 0.001
ADRB3	0.01	115	0.916
EIF4EBP1	-0.10	117	0.283
ASH2L	0.78	97	< 0.001
LSM1	0.71	88	< 0.001
BAG4	0.63	131	< 0.001
DDHD2	0.63	150	< 0.001
PPAPDC1B	0.64	150	< 0.001
NSD3	0.88	152	< 0.001
LETM2	0.61	26	< 0.001
FGFR1	0.73	152	< 0.001
TACC1	0.07	151	0.393

Bold; genes with statistically significant positive correlation between DNA copy number and mRNA expression, r: pearsons correlation coefficient, n: number of samples used for correlation calculations, p-value was used for determination of statistical significance, $p < 0.05$ was considered statistically significant. Data for mRNA expression of the genes *LOC157860, GOT1L1* and *STAR* were not available.

Statistically significant positive correlation was observed for 13 of the 17 genes. Genes that do not show positive correlation are *GPR124,*

ADRB3, EIF4EBP1 and *TACC1*. Genes with positive correlation are *ZNF703, SPFH2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1, BAG4, DDHD2, PPAPDC1B, NSD3, LETM2* and *FGFR1* displaying correlation coefficient (r) between 0.53 (*ZNF702*) and 0.88 (*NSD3*) and p-value below 0.001 (Table 6).

Protein expression of genes showing positive correlation between DNA copy number and mRNA level

Protein expression is regulated through a complex translational process. Therefore it does not necessaraly mean that protein expression is elevated when mRNA expression of amplified genes is elevated. Correlation of mRNA expression and/or DNA level to protein expression in tumour cells is an indication of an important role of the gene in tumourigenesis. To test whether an increase in DNA copy number and mRNA expression resulted in increased protein expression, protein amount was analysed in breast tumours, cell lines and primary epithelial cells. Western blotting was used for detection of the proteins and Adobe Photoshop for quantification (see Materials and methods). Antibodies were available for 11 of the 13 genes showing positive correlation between DNA copy number and mRNA expression, antibodies were not available for Letm2 and Ppapdc1b. As mentioned earlier, there were no results available for DNA copy number and mRNA expression correlation calculations for LOC157860, GOT1L1 and STAR. Antibodies were not available for Loc157860 and Got1l1 but an antibody was available for Star, thus Star was included in the protein analysis. To summarize, protein expression of the 12 following genes was analyzed: *ZNF703, SPFH2, PROSC, BRF2, RAB11FIP1, ASH2L, STAR, LSM1, BAG4, DDHD2, NSD3* and *FGFR1*.

Protein expression analysis in tumour samples

Protein expression of tumour samples was analysed by Western blotting in 39 fresh frozen tumour samples. The number of samples with amplification, gain, normal and loss at 8p12-p11 were 12, 10, 10 and 7, respectively. No more tumour samples with amplification at 8p12-p11 were available. All the samples were included in both the aCGH and GEx study and thus DNA level and mRNA amount was known.

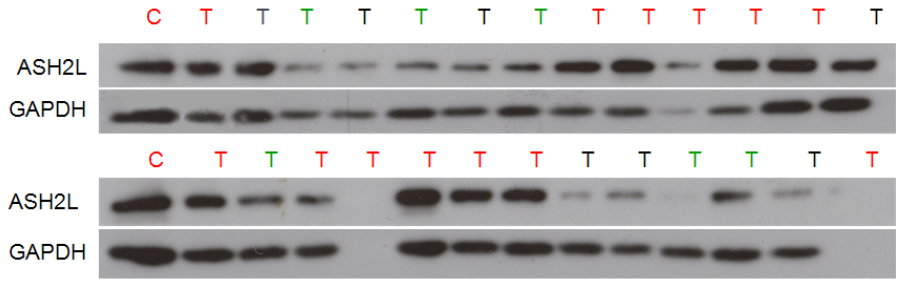


Figure 7. Ash2l protein expression in breast tumour samples with copy number imbalance and normal DNA status at 8p12-p11. This figure is an example of results from the protein analysis. Shown are results for 24 of 37 tumours analysed for Ash2l expression and one cell line (SUM52). The tumours have different DNA status at the region were *ASH2L* is located. **T** = tumour samples with amplification and/or gain of *ASH2L*, **T**= tumour samples with normal amount of *ASH2L* and **T** = tumour samples with loss of DNA at the region were *ASH2L* is located. **C** = SUM52 breast tumour cell line with amplification at 8p12-p11. No proteins were observed in two of the samples, indicating protein degradation, the two samples were exchanged for non-degraded samples in the following Western blots. Gapdh was used as a loading control.

The results from the quantification of the protein bands detected by Western blotting were used to compare the level of DNA copy number and mRNA expression. An example of a gene showing positive correlation of DNA copy number and protein expression is

*ASH2L (*Figure 7). Because of degradation of two protein samples, 37 samples were analysed for ASH2L.

Figure 7 shows protein bands from staining with Ash2l antibody for 24 tumour samples (T) of the 37 samples tested and the cell line SUM52 (C). For most of the samples with amplification/gain at the chromosomal region including the *ASH2L* gene, expression of the Ash2l protein was elevated in comparison to the protein expression in tumour samples without amplification/gain of *ASH2L*. The other 11 proteins (total of 18 protein isoforms) were analyzed in a similar manner (data not shown).

Correlation of protein expression and DNA copy number/mRNA expression of genes at 8p12-p11

The protein expression in the breast tumours was detected by Western blotting and the protein bands were quantified as described in Materials and methods. The resulting numbers were $\log_2$ transformed and correlated to $\log_2$ratio of DNA level and mRNA level to test whether enhanced DNA copy number and mRNA expression resulted in increased protein production. Pearson product-moment (r) correlation was used for correlation assessment of protein expression and DNA/mRNA level. The results were considered statistically significant if the p-value was lower than 0.05. The results are shown in Figure 8.

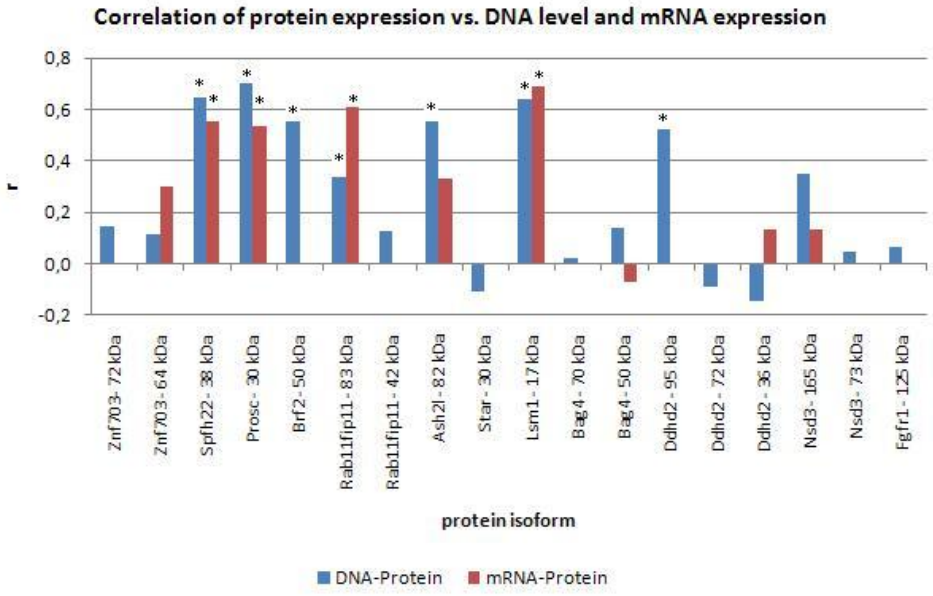


Figure 8. Correlation of protein expression vs. DNA level and mRNA expression. A histogram is depicted that shows the correlation coefficient of DNA copy number and protein expression along with the correlation coefficient (r) between mRNA levels and protein expression. The correlation coefficients between mRNA and protein expression for Star are missing due to insufficient results from the GEx study and also missing for Brf2 because too few samples had results for both mRNA level and protein level. The correlation coefficient between mRNA and expression of the protein isoforms of Znf703 72 kDa, Rab11fip1 42 kDa, Bag4 70 kDa, Ddhd2 95 kDa and 72 kDa, Nsd3 73 kDa and Fgfr1 is missing since the antibody did not detect the isoform encoded by the transcript detected by the GEx probe. On the y axis: Pearsons correlation coefficient factor (r). The genes are aligned on the x axis in the same order as found in the genome according to the UCSC Genome Browser Bulid 35 (hg17). Left to right: telomere to centromere. Blue bars: correlations between DNA copy number and protein expression, red bars: correlation between mRNA and protein expression. Statistically significant ($p < 0.05$) positive correlation is marked with * above the column.

Seven protein isoforms showed statistically significant positive correlation between its amount and DNA copy number (Spfh2, Prosc, Brf2, Rab11fip1 (83 kDa), Ash2l, Lsm1 and Ddhd2 (95 kDa). Of those, statistically significant positive correlation between mRNA and protein expression was observed for Spfh2, Prosc, Rab11fip1 (83

kDa) and Lsm1. mRNA level and protein level correlation calculations for Brf2 could not be performed because too few samples with both mRNA level and protein level data were available. Correlation between mRNA expression and Ash2l protein expression was not significant ($r = 0.332$, $p = 0.078488$), which is surprising since positive correlation between DNA level and protein expression along with correlation between DNA level and mRNA expression was observed. Correlation between mRNA level and protein amount could not be performed for Znf703 72 kDa, Rab11fip1 42 kDa, Bag4 70 kDa, Ddhd2 95 kDa and 72 kDa, Nsd3 73 kDa and Ffgr1 because the respective GEx probes did not detect the mRNA transcript encoding these isoforms. Correlation of DNA level and protein level was not observed for Znf703 (72 kDa and 64), Rab11fip1 (42 kDa), Star, Bag4 (70 kDa and 50 kDa), Ddhd2 (72 kDa and 36 kDa), Nsd3 (73 kDa and 165 kDa) and Fgfr1. Correlation was not observed between mRNA and protein level of Znf703 (64 kDa), Ash2l, Bag4 (50 kDa), Ddhd2 (36 kDa), NSD3 (165 kDa) (Figure 8).

Protein expression in primary epithelial cells and breast cell lines

Protein expression of the genes located at the 8p12-p11 amplified region was analyzed in 7 cell lines and two types of primary epithelial cells in a similar manner as in the breast tumour samples. The primary epithelial cell types (myo- and luminal-epithelial cells, MEP and LEP) were obtained from reduction mammoplasty and are assumed to have normal DNA level at 8p12-p11. Four of the cell lines have amplification and/or gain (SUM52, SUM44, HCC1500 and CaMa-1),

2 of the cell lines have loss (MCF-7 and MDA-MB-231) and one has normal (MCF-10A) DNA level at 8p12-p11, according to published data (Table 7). Protein expression of the cell lines with amplification was compared to the protein expression of the cell lines and the primary epithelial cells with loss and normal DNA status.

Table 7 DNA copy number status at 8p12-p11 in breast cell lines according to published data.

Cell line name	DNA status at the minimal region of amplification at 8p12-p11	Reference
MCF-10	normal	Jonsson et al, 2007
MCF-7	loss	Pole et al, 2006
	loss	Jonsson et al, 2007
	loss/normal	Cooke et al 2008
MDA-MB-231	loss	Jonsson et al, 2007
SUM52	amplification	Pole et al, 2006
	amplification	Cooke et al 2008
	amplification	Yang et al 2004
	amplification	Ray et al , 2004
	amplification	Yang et al, 2006
	amplification	Garcia et al, 2005
SUM44	amplification	Pole et al, 2006
	amplification	Paterson et al, 2007
	amplification	Cooke et al 2008
	amplification	Ray et al , 2004
	amplification	Yang et al, 2006
	amplification	Garcia et al, 2005
HCC1500	amplification/gain	Pole et al, 2006
	amplification/normal	Paterson et al, 2007
	amplification/gain	Cooke et al 2008
	amplification/gain	Garcia et al, 2005
CaMa-1	amplification	Pole et al, 2006
	amplification	Cooke et al 2008

Protein expression in the primary cells and cell lines were detected by Western blotting and quantified as described in materials and methods. The data are depicted in Figure 9 and Figure 10 where a bar graph for each protein isoform allows for comparison between cell lines. Note that the scaling on the Y-axis is discordant across proteins. It represents the level of each of the proteins isoforms, measured by quantification of the protein bands observed by Western blotting. The intensity of the bands depends on many factors in the Western blotting process and development of the films but the proportions between bands should be constant. Therefore, the numbers on the y-axis are unit-less and only represent differences in protein level between primary epithelial cells and cell lines and should be used for internal comparison only.

Figure 9. Protein expression in cell lines and primary epithelial cells. Expression of 9 protein isoforms in 7 cell lines, 1 immortalized cell line (MCF10A) and 2 types of primary epithelial cells (myo- and luminal epithelial cells, MEP and LEP). Cell lines and primary cells types are aligned on the x axis and the protein levels are shown on the y axis. It should be noted that the scaling can only be used for internal comparison between cell lines for each protein. The proteins are expressed from genes located at the 8p12-p11 amplified area. DNA status of the cell lines at 8p12-p11; HCC1500, CaMa-1, SUM44 and SUM52 have amplification, MCF7 and MDA-MB-231 have loss, MCF10A is normal according to published data (Table 7). MEP and LEP are assumed to be normal. The reported size of the isoforms is the one observed on the Western blots.The columns are color coded according to their DNA status at 8p12-p11. Red: amplification, yellow: normal, green: loss, na: data not available.

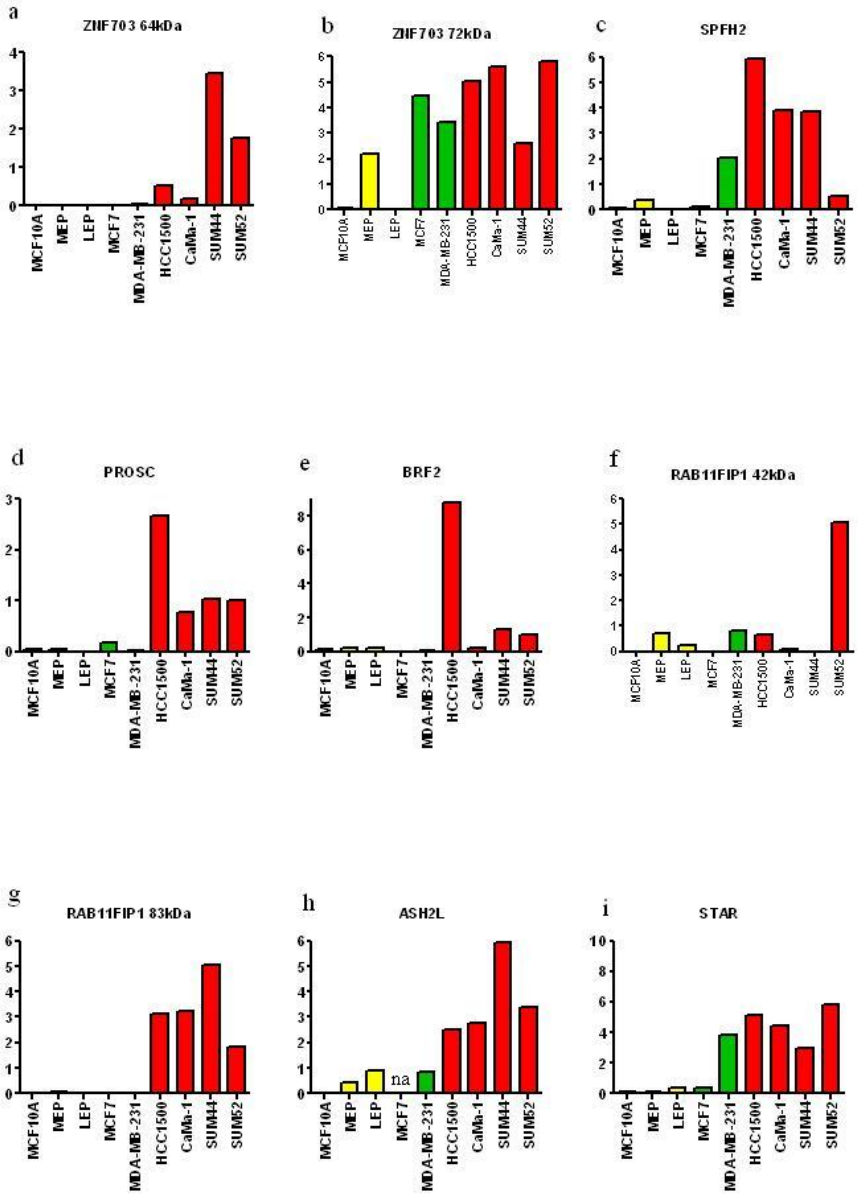
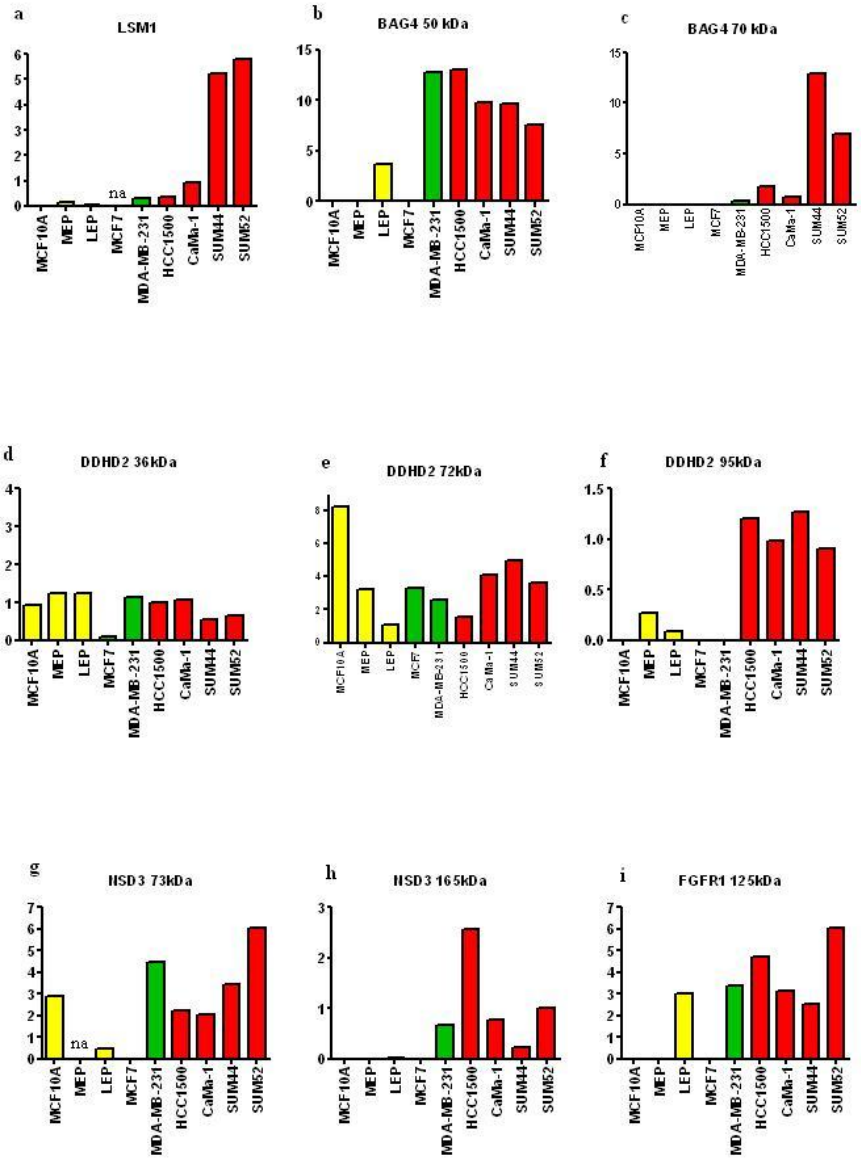

a
ZNF703 64kDa
b
ZNF703 72kDa
c
SPFH2
d
PROSC
e
BRF2
f
RAB11FIP1 42kDa
g
RAB11FIP1 83kDa
h
ASH2L
na
i
STAR
MCF10A
MEP
LEP
MCF7
MDA-MB-231
HCC1500
CaMa-1
SUM44
SUM52

Figure 10. Protein expression in cell lines and primary epithelial cells. Expression of 9 protein isoforms in 7 cell lines, 1 immortalized cell line (MCF10A) and 2 types of primary epithelial cells (myo- and luminal epithelial cells, MEP and LEP). Cell lines and primary cells types are aligned on the x axis and the protein levels are shown on the y axis. It should be noted that the scaling can only be used for internal comparison between cell lines for each protein. The proteins are expressed from genes located at the 8p12-p11 amplified area. DNA status of the cell lines at 8p12-p11; HCC1500, CaMa-1, SUM44 and SUM52 have amplification, MCF7 and MDA-MB-231 have loss, MCF10A is normal according to published data (Table 7). MEP and LEP are assumed to be normal. The reported size of the isoforms is the one observed on the Western blots.The columns are color coded according to their DNA status at 8p12-p11. Red: amplification, yellow: normal, green: loss, na: data not available.


a
LSM1
na
b
BAG4 50 kDa
c
BAG4 70 kDa
d
DDHD2 36kDa
e
DDHD2 72kDa
f
DDHD2 95kDa
g
NSD3 73kDa
na
h
NSD3 165kDa
i
FGFR1 125kDa
MCF10A
MEP
LEP
MCF7
MDA-MB-231
HCC1500
CaMa-1
SUM44
SUM52

In the following paragraphs amplification, gain, normal and loss refer only to the DNA status at 8p12-p11 in the cell lines as described in Table 7 for the cell lines. The DNA status at 8p12-p11 in MEP and LEP cells is assumed to be normal.

Expression of the 64 kDa Znf703 protein was not observed in MCF10A, MEP and LEP (normal) or in MCF7 (loss) whereas very low expression was observed in MDA-MB-231 (loss). Expression was observed in all cell lines with amplification (Figure 9a).

Expression of the 72 kDa isoform of Znf703 was high in all the cell lines with the 8p12-p11 amplified region, the ones with loss at the region (MCF7 and MDA-MB-231) and in MEP (normal). Expression was not detected in LEP and it was very low in MCF10A (both normal) (Figure 9b).

Protein expression of Spfh2 was high in three of the four cell lines with amplification. But expression was low in SUM52 (amplification), similar to the expression in MEP (normal). Low expression was observed in LEP and MCF10A (both normal) as well as in MCF-7 (loss). MDA-MB-231 (loss) expressed intermediate levels of Spfh2 higher than one of the cell lines with amplification (SUM52) (Figure 9c).

Prosc expression was higher in the cell lines with amplification than in all the cells and cell lines with normal or loss at 8p12-p11 (Figure 9d).

Brf2 expression was low in all cell lines and primary cells tested, except HCC1500 (amplification). Expression in cells and cell lines

with loss or normal DNA status was lower than in cell lines with amplification with the exception of CaMa-1. Brf2 expression in CaMa-1 was similar to the expression in MEP, LEP and MCF10A (all normal). Expression was not observed in MCF7 (loss) (Figure 9e).

Expression of the 42 kDa isoform of Rab11fip1 was highest in SUM52 (amplification). Similar amount of the protein was observed in MEP (normal), MDA-MB-231 (loss) and HCC1500 (amplification). Expression was low in LEP (normal) and CaMa-1 (amplification) and below detectable levels in MCF10A (normal), MCF7 (loss) and SUM44 (amplification) (Figure 9f)

High expression of the 83 kDa isoform of Rab11fip1 was observed in all four cell lines with amplification and not observed in either MCF10A, LEP (both normal) or MCF7 and MDA-MB-231 (both with loss). Very low expression was observed in MEP (normal) (Figure 9g).

Ash2l expression was highest in the cell lines with amplification. No protein expression was observed in MCF10A (normal) or MCF7 (loss), while expression in MEP and LEP (both normal) was similar to the expression in MDA-MB-231 (loss) (Figure 9h).

Low expression of Star was observed in all three cell lines with normal DNA status and also in MCF7 (loss). In comparison, Star expression was high in all the cell lines with amplification as well as in MDA-MB-231 (loss) (Figure 9i).

High expression of LSM1 was observed in SUM44 and SUM52, both with amplification compared to the low expression in all the primary cells and cell lines with loss and normal DNA level (MCF10A, MEP, LEP and MDA-MB-231). Expression was low in two of the cell lines with amplification (HCC1500 and CaMa-1). Data was not available for MCF7 (Figure 10a)

High expression of the 50 kDa isoform of Bag4 was observed for all the cell lines with amplification. Equally high was the expression in MDA-MB-231 (loss). No expression was observed of the 50 kDa isoform of Bag4 was in MCF10A, MEP (both normal) and MCF7 (loss) (Figure 10b).

Expression of the 70 kDa isoform of Bag4 was highest in all the cell lines with amplification, although highest in SUM52 and SUM44. Low or no expression was detected in all the cell lines and primary cell with loss and normal (Figure 10c).

Expression of the 36 kDa isoform of Ddhd2 was similar in all the cell lines with amplification, normal and loss at 8p12-p11, except in MCF7 which was the lowest (Figure 10d).

Expression of the 72 kDa isoform of Ddhd2 was fairly similar in cells and cell lines with amplification, normal and loss at 8p12-p11. However, expression was highest in MCF10A (normal) (Figure 10e).

High expression of the 95 kDa isoform of Ddhd2 was observed in all four cell lines with. Expression was not observed in MCF10A (normal), MCF7 and MDA-MB-231 (both with loss) (Figure 10f).

The highest expression of the 73 kDa isoform of Nsd3 was observed in one of the cell lines with amplification (SUM52), the second highest was observed in a cell line with loss at 8p12-p11 (MDA-MB-231). Expression was similar in MCF10A (normal) and HCC1500, CaMa-1 and SUM44 (all with amp), no expression was observed in MCF-7 (loss) and low in LEP (normal). Data were not available for MEP (normal) (Figure 10g).

Expression of the 165 kDa isoform of Nsd3 was not observed in any of the normal primary cells and cell lines or in one of the cell line with loss. However, expression in MDA-MB-231 (loss) was similar to the expression of CaMa-1 (amp). The highest expression was observed in one of the cell lines with amplification (HCC1500) while low levels were detected in SUM44 (amp) and intermediate levels in SUM52 (amp) (Figure 10h).

Expression of Fgfr1 was similar in LEP (normal) and MDA-MB-231 (loss), slightly exceeding the levels in SUM44 (amp). No protein expression was observed in MCF10A, MEP (both normal) and MCF7 (loss). The highest expression was observed in SUM52 and HCC1500 (both with amplification) (Figure 10i).

Taken together, 11 protein isoforms were more highly expressed in 3 of the 4 cell lines with amplification than in the cell lines with normal

or loss at 8p12-p11 and in primary epithelial cells (normal) (Znf703 64 kDa, Spfh2, Prosc, Brf2, Rab11fip1 (83kDa), Ash2l, Star, Lsm1, Bag4 (70 kDa), Ddhd2 95 kDa, and Nsd3 (165 kDa). Seven protein isoforms were expressed in similar amounts in most of the cell lines and primary cells (Znf703 72 kDa, Rab11fip1 42 kDa, Bag 50 kDa, Ddhd2 36 kDa and 72 kDa, Nsd3 73 kDa, Fgfr1 125 kDa).

Seven genes are the most likely candidate target genes of the 8p12-p11 amplification in breast cancer

Seven genes meet all the criteria set in this study (Table 8) and are thus the most likely candidate target genes of the 8p12-p11 amplified region in breast tumours. Those genes are *SPFH2*, *RAB11FIP1*, *PROSC, BRF2, ASH2L, LSM1* and *DDHD2*.

The seven most likely candidate target genes at the 8p12-p11 amplified region are represented in bold. LOC157860 and *GOT1L1* were not included in the GEx analysis. Sufficient data were not available for correlation calculations of mRNA and both DNA level and protein amount for *STAR*. Not enough data were available for correlation calculations between mRNA level an protein level of Brf2 because few samples had data for both. Correlation calculations between mRNA level and protein amount could not be performed for the protein isoforms; Znf703 72 kDa, Rab11fip1 42 kDa, Bag 70 kDa, Ddhd2 72 kDa and 95 kDa, 73 kDa Nsd3 and Fgfr1 because the probe in GEx did not identify the same isoform as the antibody detected. Reported sizes of the proteins (kDa) are as observed by Western blotting. ab.na: antibody not available, na; data not available, +: tested positive, 0: tested negative.

Table 8. Summary of the results obtained from the four analyses done in the study.

Gene symbol	Size (kDa)	DNA-mRNA correlation	DNA-protein correlation	mRNA-protein correlation	cell lines
ZNF703	64	+	0	0	+
	72		0	na	0
SPFH2	38	+	+	+	+
PROSC	30	+	+	+	+
GPR124		0			
BRF2	50	+	+	na	+
RAB11FIP1	83	+	+	+	+
	42		0	na	?
ADRB3		0			
EIF4EBP1		0			
ASH2L	82	+	+	0	+
STAR	30		0	0	+
LSM1	17	+	+	+	+
BAG4	50	+	0	0	0
	70		0	na	+
DDHD2	72		0	na	0
	36	+	0	0	0
	95		+	na	+
PPAPDC1B		+	ab na		
NSD3	73		0	na	0
	165	+	0	0	+
LETM2		+	ab na		
FGFR1	125	+	0	na	0
TACC1		0			

Discussion

In this study, the minimal region of amplification at 8p12-p11 was identified as a 1.6 Mb region encompassing 20 known protein-coding genes. Thirteen of the 20 genes showed positive correlation between DNA copy number and mRNA expression. Protein analyses revealed 7 genes as the most likely target genes of the amplified region.

The minimal region of amplification at 8p12-p11 is 1.6 Mb in size

The minimal region most frequently amplified at 8p12-p11 was defined as a 1.6 Mb region mapping at 37.3 – 38.9 Mb. Fifteen of the 40 samples (38%) with amplification at 8p have decrease of DNA material telomeric to the 8p12-p11 amplified region (Figure 6). That could be an indication of the amplified region being composed of double minutes formed by double strand breaks and thus loss of the telomeric region (Coquelle et al., 1997).

Mapping of the 8p12-p11 amplification has been performed in a number of studies using aCGH on primary breast tumours and breast cancer cell lines. The number of samples and/or cell lines varied between studies from only including 5 cell lines (Prentice et al., 2005) to including 152 breast tumours and 21 breast cell lines (Bernard-Pierrot et al., 2008). The resolution of the arrays varied also, from 0.08 Mb (Prentice et al., 2005) to 1.5 Mb (Garcia et al., 2005). Two studies identified four amplified regions residing at 8p12-p11, defined as A1 through A4 (Gelsi-Boyer et al., 2005; Kwek et al., 2009). However, two studies define only one amplified region, which

corresponds to A1 (Garcia et al., 2005) (Prentice et al., 2005). Yet another study identified a single 3.2 Mb minimal region of amplification which overlaps with the amplified regions A1-A3 (Bernard-Pierrot et al., 2008). All of these studies point to the region represented as A1 as the region most often amplified although boundaries vary between studies. There is more disagreement on the amplified regions defined as A2-A4. This inconsistency could partly be explained by small sample size, different sample pool and difference in methods applied. For example two of the studies used arrays not covering A4 (Garcia et al., 2005; Pole et al., 2006).

The minimal region of amplification identified in this study is 1.6 Mb in size corresponding to the centromeric part of A1 and telomeric part of A2. The 1.6 Mb region could possibly include two separate amplification targets, based on the heterogeneity of the boundaries of smaller amplified regions within this region (see samples S10, S14, S21, S23, S25, S28, S31, S34, S36 and S39) (Figure 6). These two smaller amplified regions also correspond to parts of A1 and A2. Thus, the results from mapping the 8p12-p11 amplified region in the Icelandic breast tumours are in agreement with published data.

Twenty known protein-coding genes are located within the 1.6 Mb region analysed in this study. The most telomeric protein-coding gene is located at 37.67 Mb. Thus, telomerically within the 1.6 Mb minimal region of amplification was a 370 kb region with no known protein coding genes, according to the UCSC Human Genome Browser Build 35 (hg17). There are two non-coding RNAs (ncRNA) found within

this region, according to Ensembl release 56. One of them is a U6 small nuclear RNA (snRNA) that binds other snRNAs and proteins that form the spliceosome, which is involved in RNA splicing of pre-mRNA (Stevens et al., 2001). The other is a snRNA pseudogene (AC137579.1). Neither of the two ncRNAs has been linked to tumourigenesis. Expressed sequence tags (ESTs) are also found at the region according to UCSC (UCSC Genome Browser Build 35 (hg17)) they are not included in this study.

Four criteria were set for a gene to be considered a candidate target gene of the amplified region. Correlation between DNA copy number and mRNA expression had to be positive, correlation between protein level and mRNA expression had to be observed as well as correlation between protein expression and DNA copy number. Also, the genes had to express higher levels of the protein in 3 of 4 breast cancer cell lines with amplification than in cell lines and primary epithelial cells without the amplification.

8p12-p11 is amplified in 18% of breast tumours

In this study the most common amplified regions in breast tumours, according to the aCGH data, were found at 17p12, 8p12 and 11q13, which is in accordance with previously published data (Forozan et al., 2000; Ethier, 2003; Letessier et al., 2006; Bernard-Pierrot et al., 2008). However, the frequency of the amplification varies. That could be explained by the different methods applied in differently composed sample pools and the lack of a definition of the term gain and amplification.

Co-amplification of 8p12 and 11q13 have been reported, where amplification of *FGFR1* (8p12) and *CCND1* (11q13) have been linked to reduced patient survival (Cuny et al., 2000; Kwek et al., 2009). *CCND1* is found amplified in around 40% of tumours with amplification of *FGFR1* (Courjal et al., 1997). According to aCGH data for the Icelandic tumour samples, 30% of samples with amplification of *FGFR1* are also amplified at *CCND1* (data not shown).

Tumour samples are rarely homogenous. Immune cells, fibroblasts and endothelial cells, for example, are often found amongst malignant cells in the tumour. The 8p12-p11 amplification is a somatic aberration and therefore not present in those non-tumour cells found within the tumour. Thus, when the DNA was extracted from the tumour samples for hybridization on the CGH arrays, the DNA from the non-tumour cells dilute the results of the DNA aberrations found in the tumour cells. Assuming that non-tumour cells are around 20% of the tumour samples (average figure according to personal information from Rósa Björk Barkardóttir), 20% dilution can be expected on the results obtained from the array-CGH. Thus, 2 extra copies of a certain chromosomal region is equal to a ratio of 1.6 (comparing tumour DNA to normal DNA). In this study, two or more extra copies of a certain chromosomal region were considered as amplification. Thus, samples with ratio ≥ 1.6 according to the aCGH data were classified as amplified. Likewise ratio of 3 or more equals to 3.8 or more extra copies. Consensus is needed regarding the definition of gain and amplification of DNA. In published studies an

amplification has been defined from a ratio of 1.2 and up to 2 (Ray et al., 2004; Garcia et al., 2005; Gelsi-Boyer et al., 2005; Yang et al., 2006; Bernard-Pierrot et al., 2008; Kwek et al., 2009).

According to the definition above, the amplification of the 8p12-p11 chromosomal region was observed in 18% of breast tumours. In the original studies on the 8p12-p11 region, amplification was estimated in 10-15% of breast tumours using Southern blotting (Theillet et al., 1993; Courjal et al., 1997; Adelaide et al., 1998). More recent studies using aCGH at high resolution observe the amplification in 13-25% of breast tumours (Garcia et al., 2005; Prentice et al., 2005; Letessier et al., 2006; Yang et al., 2006). A frequency around 7% has recently been published (Bernard-Pierrot et al., 2008; Kwek et al., 2009). Therefore, the frequency in the Icelandic samples falls within the middle of the reported frequency range 7 – 25%. This frequency variation can most likely be explained by the difference in array resolutions, the different sample pools and in inconsistent definition of amplification.

Thirteen genes show positive correlation between DNA copy number and mRNA level

Candidate oncogenes should be overexpressed when amplified, although it should be noted that overexpression may also occur by other means. Of the 17 genes analysed for correlation of DNA copy number and mRNA expression, 13 showed positive correlation (*ZNF703, SPFH2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1, BAG4, DDHD2, PPAPDC1B, NSD3, LETM2* and *FGFR1*) (Table 6).

Positive correlation between DNA copy number and mRNA expression has been observed previously for all the 13 genes (Garcia et al., 2005; Gelsi-Boyer et al., 2005; Yang et al., 2006; Bernard-Pierrot et al., 2008; Kwek et al., 2009). Positive correlation was not observed for 4 of the tested genes (*GPR124, ADRB3, EIF4EBP1* and *TACC1*). Two other studies have tested the correlation between *GPR124* and *ADRB3* and neither of them observed positive correlation (Bernard-Pierrot et al., 2008; Kwek et al., 2009). However, two studies have tested the correlation between DNA copy number and mRNA expression of *EIF4EBP1* and both of them observed positive correlation (Yang et al., 2006; Bernard-Pierrot et al., 2008). In this study the correlation between DNA copy number and mRNA expression was $r = -0.10$ ($p = 0.283$) for 117 primary breast tumour samples while Yang et al. used a total of 33 samples (28 tumours and 5 cell lines) for their correlation calculations and according to Spearman correlation coefficient (ρ) the correlation was positive ($\rho = 0.54$, $p = 0.005$). Also, Bernard-Pierrot et al. observed correlation of $r = 0.677$ and $r = 0.578$ ($p < 0.05$) in 18 cell lines and 128 tumours, respectively. The contradiction between the studies performed by Yang et al. and Bernard-Pierrot et al. can unlikely explained by difference in resolution of the arrays because Yang et al. used an array with resolution of 43 kb over the genome while Bernard-Pierrot et al. used an array with resolution of 1000 kb over the genome. In this study the resolution of the array was 80 kb over the genome. Differently composed sample pools could most likely explain this difference in results. The lack of positive correlation in the Icelandic

samples was convincing and thus *EIF4EBP1* was not included in further tests. One study tested the correlation between DNA copy number and mRNA expression of *TACC1* and observed positive correlation for tumours but not for cell lines (Kwek et al., 2009). Kwek et al, correlate their DNA copy number results to three GEx probes each binding one different transcript and obtain positive correlation between two probes but not the third one. In this study, the results for mRNA level of *TACC1* were obtained from a probe binding to 5 different transcripts. It can therefore not be excluded that one or more of TACC1 transcripts correlate with DNA amplification status. Information on the mRNAs transcribed from the genes *LOC157860* and *GOT1L1* are lacking; therefore probes were not included on the GEx arrays. Also lacking in the GEx array data available in this study were expression data for *STAR*. Probes for STAR were included on the GEx arrays but the quality of the data was not good enough to be used. Two studies have analysed the correlation of DNA copy number and expression of *STAR* and neither of them observed positive correlation (Bernard-Pierrot et al., 2008; Kwek et al., 2009). Thus, the results for the 13 genes showing positive correlation between DNA copy number and mRNA expression confirm previously published data. It should be noted that the quantity of mRNA does not provide a measurement of the level of protein expression. Proteins are in most cases the functional units of genes. Altered level of a certain protein could have adverse effect on a cell. Thus, measuring the level of protein expression of a certain gene in tumour samples and correlating

with DNA status could provide positive indication of the possible tumourigenicity of the gene in question.

Protein analysis in breast tumours implied seven genes as possible target genes of the 8p12-p11 amplification

Correlation between DNA level and protein expression of the genes at the 8p12-p11 amplified region in breast tumours has not been observed in any studies published to date. Many studies have observed correlation between DNA level and mRNA expression of the genes in cell lines and tumour cells for indication of target gene of the amplified region. But, a complex translational and post-translational process ultimately determines the specific quantities of the proteins produced. It has been shown that there is often poor correlation between mRNA and protein abundance (Anderson &Seilhamer, 1997; Gygi et al., 1999)

To assess whether the observed overexpression of mRNA resulted in overexpression of the protein the correlation of DNA amplification, mRNA level and protein amount was analysed. Positive correlation is an indication of the gene being a target gene of the amplification, thus possibly having significance in the tumourigenicity. Four genes showed correlation between both DNA level and protein amount and between mRNA and protein amount *(SPFH2, PROSC, RAB11FIP1* and *LSM1)*. Correlation between DNA level and protein amount of Brf2 was observed, but data were lacking for correlation of mRNA level and protein expression of Brf2 because too few samples had data for both mRNA level and protein amount. Data for correlation of the

level of transcripts coding for the 95 kDa Ddhd2 to both DNA level and protein amount was lacking since the probe in GEx identified the transcript coding for the 36 kDa protein, but positive correlation was observed between DNA level and amount of the 95 kDa isoform of Ddhd2. Correlation between mRNA level and protein amount of *ASH2L* was not observed which was surprising since correlation was observed between DNA level and mRNA level and also for DNA level and protein amount. The probe in GEx used to detect mRNA level recognises two transcripts coded by *ASH2L* but the antibody used to detect protein amount only recognises the protein coded by one of these two transcripts. Therefore if the other transcript (the one coding for the protein that the antibody does not recognise) is expressed unevenly between tumours it could have effect on the correlation obtained between mRNA level and the protein. Therefore, both *ASH2L* and *DDHD2* were considered possible target genes. Taken together, seven proteins showed positive correlation to DNA level and/or mRNA level (Spfh2, Prosc, Brf2, Rab11fip1 (83 kDa isoform), Ash2l, Lsm1 and Ddhd2 (95 kDa isoform)).

Protein analysis in breast cell lines

Protein expression of the genes identified as overexpressed when amplified was also analysed in cell lines for support and comparison of the results obtained from the tumours. One indication of an amplified gene being an oncogene is if it expresses proteins at higher level in cancerous cells with the amplification than in cancerous cells without the amplification and non-cancerous cells. Eleven protein

isoforms were expressed at higher levels in at least 3 of 4 of the cell lines with amplification at 8p12-p11 than in cell lines with normal status or loss at the region and normal cultured epithelial cells (Znf703 64 kDa, Spfh2, Prosc, Brf2, Rab11fip1 83 kDa, Ash2l, Star, Lsm1, Bag4 70 kDa, Ddhd2 95 kDa and Nsd3 165kDa). There could be a microdeletion in any of the cell lines with amplification which was not detected by aCGH or some technical problem which lowered the protein amount detected in the Western blotting and/or the quantification of the protein amount. Therefore it was considered valid if expression was higher in at least 3 of the 4 cell lines harbouring amplification at 8p12-p11. The protein expression in the cell lines and primary cells suggests that seven isoforms are not significant in tumourigenesis (Znf703 72 kDa, Rab11fip1 42 kDa, Bag 50 kDa, Ddhd2 72 kDa, Ddhd2 36 kDa, Nsd3 73 kDa and Fgfr1 125kDa). Most likely there is some other protein expression regulation mechanism than gene amplification behind the expression of these protein isoforms.

Four of the protein isoforms with higher expression in at least 3 of the 4 cell lines with amplification (Znf703 64 kDa, Star, Bag4 70 kDa, and Nsd3 165 kDa) did not show correlation between protein level and DNA/mRNA level in primary breast tumours. This could be explained by difference in the protein expression regulation between cell lines and tissue tumour cells. The regulation could be more complex in the tumour cells than in the cell lines and other factors than amplification of the gene playing a part in the protein expression.

Seven of the protein isoforms with higher expression in at least 3 of the 4 cell lines with amplification also showed correlation between DNA or mRNA level and protein level, thus supporting the results obtained from the primary breast tumours (Spfh2, Prosc, Brf2, Rab11fip1 (83 kDa), Ash2l, Lsm1, Ddhd2 (95 kDa)).

In another study, protein expression was tested in four of the same cell lines used in this study (SUM52, SUM44 (both with amplification), MCF7 (loss) and MCF10A (normal))(Kwek et al., 2009). Expression of the cell lines with amplification and loss were compared to expression MCF10A. If all of the cell lines bearing amplification expressed higher level of the protein of interest than MCF10A then the gene was considered a candidate oncogene in that study. Based on that, RAB11FIP1, ASH2L, LSM1, BAG4, NSD3 and FGFR1 but not BRF2 and GPR124 were defined as candidate oncogenes. The results are in agreement with the ones obtained in this study with the exception of Brf2 which was highly expressed in MCF10A in the study performed by Kwek et al. while no expression was observed in MCF10A in this study. Also *FGFR1* was identified as a candidate oncogene by Kwek et al. based on high expression in two cell lines with amplification at 8p12-p11 (MDA-MB-134 and SUM44), no expression was observed in SUM52. Based on the protein analysis in the cell lines in this study, *FGFR1* is not a candidate oncogene, because of higher expression of the primary luminal epithelial (LEP) cells and in one cell line with loss than in two cell lines with amplification (SUM44 and CaMa-1). Although using the same cell

line types the difference of the results for Brf2 and Fgfr1 could be explained by changes over time in the expression regulation of the cell lines. It should be pointed out that the expression of Znf703, Spfh2, Prosc, Star, and Ddhd2 was not measured in the study performed by Kwek et al. Also in that study the criteria for the gene to be considered a likely oncogene the protein had to be expressed higher in the cell lines with amplification (SUM52 and SUM44 amongst others) than in MCF10A. In this study the expression of the protein in the cell lines with amplification was compared to the expression in MCF10A and primary cells and cell lines with loss at 8p12-p11. In a comparison between this study and Kwek et al. the protein expression in SUM52 and SUM44 compared to protein expression in MCF10A (the same way as was done in Kwek et al.) revealed that expression in SUM52 and SUM44 was higher than in MCF10A of all the protein isoforms examined, except the 36 kDa and 72 kDa isoforms of Ddhd2.

Seven of the 20 protein coding genes at 8p12-p11 are candidate target genes of the amplification

Taking together, seven genes met all the criteria set in this study for a gene to be considered an oncogene they are; *SPFH2, PROSC, BRF2, RAB11FIP1, ASH2L, LSM1* and *DDHD2*. However 9 genes were identified as unlikely target genes at the region (*ZNF703*, *GPR124, ADRB3, EIF4EBP1, STAR*, *BAG4, NSD3 FGFR1* and *TACC1)* because they did not meet the criteria set in the study. Two genes were overexpressed when amplified but protein expression could not be analysed due to lack of commercially available antibodies at the time the study was performed (*LETM2* and *PPAPDC1B*).

Function of the seven candidate target genes of the amplification

The function of the seven candidates has been observed in published studies. Spfh2 along with Spfh1 form a heteroligomeric complex that selectively targets activated IP_3 receptors in the endoplasmic reticulum (ER) for elimination via the ERAD pathway (endoplasmic reticulum associated degradation) which is responsible for degradation of aberrant proteins in the ER (Vembar &Brodsky, 2008; Wang et al., 2009; Wojcikiewicz et al., 2009). IP_3 receptors form Ca^{2+} gated channels in the ER membrane and play a role in cell signalling (Foskett et al., 2007).

PROSC (proline synthetase co-transcribed homolog) is a highly conserved gene that encodes a protein likely to be in a soluble cytoplasmic fraction, but its function remains to be determined. The gene is highly homologous to putative genes of bacteria species associated with proline synthetases (Ikegawa et al., 1999).

BRF2 codes for a subunit of the RNA polymerase III transcription factor complex (TFIIIB), and binds TATA binding proteins at the TATA binding box of small nuclear ribonucleic acid (snRNA) promoters (Schramm et al., 2000; Cabart &Murphy, 2001). Deregulation of *BRF2* expression has been suggested as the main mechanism behind observed deregulation of RNA polymerase III transcription in cancer cells (Cabarcas et al., 2008).

RAB11FIP1 codes for a small GTPase proteins that play essential roles in membrane traffic regulation (Deneka et al., 2003). Alteration in expression of some of the Rab proteins could potentially be linked

to cellular transformation, for example by their influence on growth factor receptor trafficking and signalling (as reviewed in (Tang &Ng, 2009)). The Rab11fip1 proteins have been suggested to modulate integrin trafficking in ways that appear to enhance cancer cell adhesion and migration (Jin &Goldenring, 2006; Caswell et al., 2008).

Ash2L is part of histone methylation transferase complex (Hughes et al., 2004). It is found in the MLL (mixed-lineage leukemia) complex that is associated with a variety of aggressive leukemias in children and adults (Tenney &Shilatifard, 2005).

Lsm1 is a member of the highly conserved Sm-like protein family which is involved in mRNA decapping and degradation (Zaric et al., 2005) (Achsel et al., 1999). The family consist of eight proteins Lsm1-Lsm8 that form two distinct heteroheptameric rings the Lsm1-7 complex and the Lsm2-8 complex (Tharun et al., 2000) (Bouveret et al., 2000). In mammalian cells the Lsm1-7 complex is required for decapping and degradation of mRNA containing AU-rich elements and specific histone mRNA's (Stoecklin et al., 2006; Mullen &Marzluff, 2008).

DDHD2 is a member of the PA-PLA_1 family (phosphatidic acid-preferring phospholipase A1) (Nakajima et al., 2002). *DDHD2* codes for a protein bound to the membrane of the endoplasmic reticulum (ER) (Nakajima et al., 2002; Shimoi et al., 2005). Its overexpression in cultured cells causes aggregation of the ER thought to be explained by the Ddhd2 changing the shape of the lipids in the cytoplasmic

leaflet of the lipid bilayer (Nakajima et al., 2002). Three isoforms of Ddhd2 have been identified, 81, 37.6 and 30.7 kDa according to Ensembl release 57 and (Nakajima et al., 2002). Two of them (81 kDa and 37.6 kDa) were detected by Western blotting in this study and neither of them showed correlation between either DNA copy number or mRNA expression and protein expression. A fourth isoform was detected by Western blotting in this study. It is a 95 kDa isoform that has not yet been mentioned in the literature. Of the three transcripts of *DDHD2* known the biggest one is reported to encode a 85 kDa isoform, it is a possible that the 95 kDa isoform observed in this study was encoded by that transcript or the fourth transcript is produced in the cells but was not detected in the other published studies. Positive correlation was observed between DNA level and expression of this 95 kDa isoform (Figure 8). Also cell lines with amplification at 8p12-p11 express the isoform in higher amount than primary cells and cell lines with normal status or loss at 8p12-p11.

Results for the seven candidate target genes in other studies

Other studies analysing the 8p12-p11 amplification trying to identify the oncogene(s) at the region have done so using various methods. By analysing the protein expression in cell lines with amplification at 8p12-p11 and comparing to a cell line without the amplification (MCF10A), *BRF2* has been excluded and *LSM1* included as potential oncogene at 8p12-p11 (Kwek et al., 2009). In this study expression of Brf2 in MCF10A was not observed. *PROSC* has also been excluded as relevant gene of oncogenesis in tumour cells harbouring the 8p12-p11

amplification by two groups, based on it not always being overexpressed when amplified (Garcia et al., 2005; Kwek et al., 2009).

By soft agar assay, the transforming potential of a gene can be assessed. Non-malignant cells are anchor dependent and thus do not grow in soft agar. By overexpressing an oncogene in a non-malignant cell line the cells can become oncogenic resulting in anchor independency, thus the cells become able to form colonies in soft agar. One study reported the lack of transforming potential of *SPFH2*, *RAB11FIP1*, *ASH2L*, *LSM1,* and *DDHD2* when overexpressed by themselves, but when *DDHD2* was co-expressed with *MYC* colonies were observed (Kwek et al., 2009). However, another group has reported the transformation potential of *LSM1* using the same method (Yang et al., 2006; Streicher et al., 2007). Thus, the results from the studies obtained by Kwek et al. and Streicher et al. are contradictory regarding the transformation potential of *LSM1*. A possible explanation could be that Kwek et al. observed that late passages of MCF10A developed the ability to grow in soft agar. Thus, they selected MCF10A cells lacking soft agar growth capability by clonal-dilution and only infected those specially selected cells whereas Streicher et al. did not select MCF10A cells lacking the ability to grow in soft agar. One other group has tested the transformation potential of *LSM1* with the addition of testing whether its overexpression resulted in growth factor independence of the cells (Yang et al., 2006). They observed that overexpression of *LSM1* in non-malignant cells resulted in transformation and insulin

independence, but *LSM1* needed to be co-expressed with *BAG4* or *C8orf4* to gain EGF (epidermal growth factor) independency. Based on the inconsistency of results on the tumourigenic potential of *LSM1* and the fact that different methods yield different results, further research is necessary to determine whether the gene is a potential target gene of the 8p12-p11 amplification.

Another way to assess the potential of a gene being an oncogene is to knock down (e.g. using siRNA) its expression in cancer cell lines harbouring the amplification and observe whether the growth rate decreases. Three of the seven candidates have been analysed that way (*DDHD2, LSM1* and *RAB11FIP1*). Decrease in mRNA level was reported in cell lines with amplification at 8p12-p11 by inhibition with gene-specific siRNA for *DDHD2* and *LSM1* (Streicher et al., 2007) (Bernard-Pierrot et al., 2008). Although after further tests on the genes Bernard-Pierrot et al. excluded them as potential oncogenes and identified two others as likely target genes of the amplification (*NSD3* and *PPAPDC1B*). *RAB11FIP1* has been identified as a candidate oncogene of the 8p12-p11 amplified region, based on transformation potential and that knock-down inhibited colony formation, invasion and migration *in vitro* and reduced tumour formation and metastasis *in vivo*. Also, it was suggested as a potential therapeutic target (Zhang et al., 2009). However Kwek et al. reported that *RAB11FIP1* did not succeed in transforming the cells when overexpressed in non-tumourigenic cells. This difference could be explained by the clonal diffusion used by Kwek et al. and possible difference in growth medium.

Taking together the results of the functional studies on the seven most likely target genes identified in this study. Four of the seven candidate oncogenes identified in this study have been identified as unlikely oncogenes (*SPFH2*, *PROSC*, *BRF2* and *ASH2L*) and disagreement is about the tumourigenic properties of *LSM1*. However, two of the seven genes have more often been described as likely target genes of the 8p12-p11 amplification that is the genes *RAB11FIP1* and *DDHD2*.

Function of the nine unlikely candidate target genes and results from other studies

Of the nine genes identified as unlikely candidate target genes, four did not show correlation between DNA level and mRNA level (*GPR124, ADRB3, EIF4EBP1* and *TACC1*) and thus were not included in the protein analysis. They will not be discussed further in this report. The other five genes (*ZNF703, STAR, BAG4, NSD3* and *FGFR1*) showed positive correlation between DNA level and mRNA level but based on the protein analysis were not likely candidate target genes.

Even though the 5 genes were identified as unlikely candidate genes of the amplified region in this study they could potentially cooperate with other oncogenes at either 8p12-p11 or genes located elsewhere in the genome. In other studies, *ZNF703*, *BAG4* and *FGFR1* have been suggested as candidate oncogenes of the 8p12-p11 amplified region or playing a part in oncogenesis (Yang et al., 2006; Bernard-Pierrot et al., 2008; Kwek et al., 2009).

The Znf703 protein contains a zinc-finger domain, a protein motif for nucleic acid recognition and has been found to repress transcription in zebrafish (Ladomery &Dellaire, 2002; Ray et al., 2004; Nakamura et al., 2008). One study identified *ZNF703* as functionally cooperating with major pathways in oncogenesis (Kwek et al., 2009).

StAR (steroidogenic acute regulatory protein) is a transport protein that regulates cholesterol transfer within the inner membrane space of the mitochondria of cells for biosynthesis of steroids (Stocco &Clark, 1996) (Lin et al., 1995). The gene has not been identified as candidate target gene of the amplified region by any of the studies published to date.

Bag4 binds to tumour necrosis factor receptor 1 (TNFR1) and keeps it in an inactive, monomer state (Forozan et al., 2000). Active TNFR1, either by TNF binding or dimerization, initiates a pathway leading to activities such as apoptosis and activation of transcription (Tartaglia et al., 1993) (Hsu et al., 1995). *BAG4* has been proposed as breast cancer oncogene based on its transforming potential and insulin independence when overexpressed by itself and EGF independence when co-expressed with either *LSM1* or *C8orf4* (Yang et al., 2006), while it has also been found identified as non-transforming (Bernard-Pierrot et al., 2008; Kwek et al., 2009) .

The function of the Nsd3 protein has not been described yet. *NSD3* is a member of the NSD family along with *NSD1* and *NSD2*. Proteins in the Nsd family contain conserved domains believed to be involved in transcriptional regulation and modulation of chromatin structure

(Angrand et al., 2001). It has been shown to play a role in regulation of survival and transformation (Bernard-Pierrot et al., 2008).

FGFR1 encodes for a protein of the fibroblast growth factor receptor (FGFR) family. It is a tyrosine kinase receptor which binds fibroblast growth factors mediating mitogenesis and differentiation. It was at first considered the prime candidate target gene of the 8p12-p11 amplified region (Adelaide et al., 1998; Ugolini et al., 1999). In recent studies it has been shown that *FGFR1* is not always included in the amplified region and it is not always overexpressed when amplified, therefore is has been suggested not to be one of the main target gene of the amplification (Ray et al., 2004; Garcia et al., 2005; Gelsi-Boyer et al., 2005; Prentice et al., 2005; Pole et al., 2006). Inhibition of *FGFR1* expression has not been shown to slow the proliferation of cell lines with the 8p12-p11 amplification (Ray et al., 2004; Reis-Filho et al., 2006b). Also, one study reported *FGFR1* as being non-transforming a cells when overexpressed by itself in non-tumourigenic cell line, but when expressed together with *MYC* the cells transformed (Kwek et al., 2009).

Genes not included in the protein analysis

Two genes showed positive correlation between DNA level and mRNA level but antibodies were not available at the time the study was performed (*PPAPDC1B* and *LETM2*). Function of the protein encoded by *PPAPDC1B* (Phosphatidic acid phosphatise type 2 domain containing 1B) is not well known but there are indications that it is a transmembrane protein functioning as a phosphatase of

phosphatidic acid (PA) in mammalian plasma membrane (Takeuchi et al., 2007). Its potential as a therapeutic target was shown by positive transformation tests and that silencing of *PPAPDC1B* in breast tumour xenografts lowered tumour volume by 80% (Bernard-Pierrot et al., 2008). The second gene, *LETM2,* codes for a leucine zipper-EF-hand containing transmembrane protein 2 (according to (Entrez-Gene)). One other study has also observed positive correlation between DNA level and mRNA level (Yang et al., 2006). The gene has not been included in protein analysis or functional studies published to date.

Future steps

It is important to continue to study the 8p12-p11 amplified region to narrow down the possible target genes. With this study the number of candidate target genes is down to seven. To observe whether the expression of the proteins encoded by the amplified genes are abnormal it would be informative to analyse the expression in normal tissue samples from the same patients as the tumours came from. Increased amount of proteins expressed by the amplified gene of interest in the tumour compared to the expression in the normal tissue could indicate a role in oncogenesis played by the gene. Expression of the 12 proteins analysed in breast tumours in this study was also analysed in normal samples from part of the same patients as the tumour samples. Due to degradation in some protein samples and technical problems the results could not be used for the analysis.

To obtain further information on the tumourigenicity of the genes and their role in survival of the malignant cells, a knock down assay

(siRNA) has been started in cell lines with amplification at 8p12-p11, as a part of a master's project of a student at the Department of Pathology. Also it would be informative to do a soft agar assay to assess the gene's tumourigenic potential. The genes that are able to transform non-malignant cells would be good candidates for further functional studies. It would be interesting to xenograft cells harbouring the amplified gene in an animal model to assess if tumour growth can be decreased by using siRNA specific to the gene of interest. Observing whether targeting the protein with antibody would neutralise its function and decrease the tumour growth would be informative.

Conclusions

The 8p12-p11 amplified region is one of the most frequently amplified regions in breast carcinoma and therefore likely to contain one or more oncogenes. Although a number of research groups have focused on the amplified region, the target gene(s) of the amplification has not yet been fully clarified. There are a number of indications that there is more than one oncogene located within the amplified region(s) at 8p12-p11. Also, different oncogenes could be the target in different tumour subgroups.

For these results to be utilized for further research aiming for therapeutic design it is important into take in account location and function of the proteins encoded by the amplified genes. Also, some proteins are better therapeutic targets than others (Hopkins &Groom, 2002). Further research is necessary to try to solve the target gene's

signatures at the 8p12-p11 amplified region. Better understanding of the target gene/s of the 8p12-p11 amplified region in breast cancer and the functional consequences of their overexpression could possibly enable the design of therapeutics for the subgroup of breast cancer patients bearing tumours with this somatic aberration. By identifying *SPFH2*, *PROSC*, *RAB11FIP1*, *BRF2*, *ASH2L*, *LSM1* and *DDHD2* as potential target genes of the 8p12-p11 amplified region in breast cancer one more puzzle has been added to the giant puzzle of cancer biology.

REFERENCES

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Luhrmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. Embo J 18, 5789-802.

Adelaide, J., Chaffanet, M., Imbert, A., Allione, F., Geneix, J., Popovici, C., van Alewijk, D., Trapman, J., Zeillinger, R., Borresen-Dale, A.L., Lidereau, R., Birnbaum, D. and Pebusque, M.J. (1998) Chromosome region 8p11-p21: refined mapping and molecular alterations in breast cancer. Genes Chromosomes Cancer 22, 186-99.

Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Kochli, O., Zuber, M., Dieterich, H., Mross, F., Wilber, K., Simon, R. and Sauter, G. (2004) Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 64, 8534-40.

Albertson, D.G. (2006) Gene amplification in cancer. Trends Genet 22, 447-55.

Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 18, 533-7.

Angrand, P.O., Apiou, F., Stewart, A.F., Dutrillaux, B., Losson, R. and Chambon, P. (2001) NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. Genomics 74, 79-88.

Banerjee, D., Mayer-Kuckuk, P., Capiaux, G., Budak-Alpdogan, T., Gorlick, R. and Bertino, J.R. (2002) Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. Biochim Biophys Acta 1587, 164-73.

Barcellos-Hoff, M.H. and Ravani, S.A. (2000) Irradiated mammary gland stroma promotes the expression of tumorigenic potential by unirradiated epithelial cells. Cancer Res 60, 1254-60.

Bernard-Pierrot, I., Gruel, N., Stransky, N., Vincent-Salomon, A., Reyal, F., Raynal, V., Vallot, C., Pierron, G., Radvanyi, F. and Delattre, O. (2008) Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer Res 68, 7165-75.

Bharat, A., Gao, F. and Margenthaler, J.A. (2009) Tumor characteristics and patient outcomes are similar between invasive lobular and mixed invasive ductal/lobular breast cancers but differ from pure invasive ductal breast cancers. Am J Surg 198, 516-9.

Blanco, A., Grana, B., Fachal, L., Santamarina, M., Cameselle-Teijeiro, J., Ruiz-Ponte, C., Carracedo, A. and Vega, A. (2009) Beyond BRCA1 and BRCA2 wild-type breast and/or ovarian cancer families: germline mutations in TP53 and PTEN. Clin Genet.

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. and Seraphin, B. (2000) A Sm-like protein complex that participates in mRNA degradation. Embo J 19, 1661-71.

Brauch, H. and Jordan, V.C. (2009) Targeting of tamoxifen to enhance antitumour action for the treatment and prevention of breast cancer: the 'personalised' approach? Eur J Cancer 45, 2274-83.

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

Cabart, P. and Murphy, S. (2001) BRFU, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with TATA-binding protein. J Biol Chem 276, 43056-64.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. and Croce, C.M. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A 101, 2999-3004.

Carter, P., Presta, L., Gorman, C.M., Ridgway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E. and

Shepard, H.M. (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc Natl Acad Sci U S A 89, 4285-9.

Caswell, P.T., Chan, M., Lindsay, A.J., McCaffrey, M.W., Boettiger, D. and Norman, J.C. (2008) Rab-coupling protein coordinates recycling of alpha5beta1 integrin and EGFR1 to promote cell migration in 3D microenvironments. J Cell Biol 183, 143-55.

Catteau, A., Harris, W.H., Xu, C.F. and Solomon, E. (1999) Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics. Oncogene 18, 1957-65.

Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, E., Godwin, J., Gray, R., Hicks, C., James, S., MacKinnon, E., McGale, P., McHugh, T., Peto, R., Taylor, C. and Wang, Y. (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet 366, 2087-106.

Comen, E.A. and Robson, M. Inhibition of poly(ADP)-ribose polymerase as a therapeutic strategy for breast cancer. Oncology (Williston Park) 24, 55-62.

Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. and Debatisse, M. (1997) Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. Cell 89, 215-25.

Coquelle, A., Toledo, F., Stern, S., Bieth, A. and Debatisse, M. (1998) A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. Mol Cell 2, 259-65.

Courjal, F., Cuny, M., Simony-Lafontaine, J., Louason, G., Speiser, P., Zeillinger, R., Rodriguez, C. and Theillet, C. (1997) Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. Cancer Res 57, 4360-7.

Cuny, M., Kramar, A., Courjal, F., Johannsdottir, V., Iacopetta, B., Fontaine, H., Grenier, J., Culine, S. and Theillet, C. (2000) 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 60, 1077-83.

Deneka, M., Neeft, M. and van der Sluijs, P. (2003) Regulation of membrane transport by rab GTPases. Crit Rev Biochem Mol Biol 38, 121-42.

Dieci, G., Preti, M. and Montanini, B. (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics 94, 83-8.

EBCTCG, E.B.C.T.C.G. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. In: Lancet, Vol. 365, p. 1687-717.

Elsheikh, S., Green, A.R., Aleskandarany, M.A., Grainge, M., Paish, C.E., Lambros, M.B., Reis-Filho, J.S. and Ellis, I.O. (2008) CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. Breast Cancer Res Treat 109, 325-35.

Entrez-Gene. http://www.ncbi.nlm.nih.gov/gene/137994?ordinalpos=2&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum. In, Vol. 2010.

Ethier, S.P. (2003) Identifying and validating causal genetic alterations in human breast cancer. Breast Cancer Res Treat 78, 285-7.

Fisher, B., Anderson, S., Bryant, J., Margolese, R.G., Deutsch, M., Fisher, E.R., Jeong, J.H. and Wolmark, N. (2002) Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 347, 1233-41.

Forozan, F., Mahlamaki, E.H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G.C., Ethier, S.P., Kallioniemi, A. and Kallioniemi, O.P. (2000) Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. Cancer Res 60, 4519-25.

Foskett, J.K., White, C., Cheung, K.H. and Mak, D.O. (2007) Inositol trisphosphate receptor Ca2+ release channels. Physiol Rev 87, 593-658.

Foster, S.A., Wong, D.J., Barrett, M.T. and Galloway, D.A. (1998) Inactivation of p16 in human mammary epithelial cells by CpG island methylation. Mol Cell Biol 18, 1793-801.

Garcia, M.J., Pole, J.C., Chin, S.F., Teschendorff, A., Naderi, A., Ozdag, H., Vias, M., Kranjac, T., Subkhankulova, T., Paish, C., Ellis, I., Brenton, J.D., Edwards, P.A. and Caldas, C. (2005) A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. Oncogene 24, 5235-45.

Gelsi-Boyer, V., Orsetti, B., Cervera, N., Finetti, P., Sircoulomb, F., Rouge, C., Lasorsa, L., Letessier, A., Ginestier, C., Monville, F., Esteyries, S., Adelaide, J., Esterni, B., Henry, C., Ethier, S.P., Bibeau, F., Mozziconacci, M.J., Charafe-Jauffret, E., Jacquemier, J., Bertucci, F., Birnbaum, D., Theillet, C. and Chaffanet, M. (2005) Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 3, 655-67.

Goldenberg, M.M. (1999) Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. Clin Ther 21, 309-18.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M.J. and Petersen, O.W. (2002) Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J Cell Sci 115, 39-50.

Gudmundsson, J., Johannesdottir, G., Arason, A., Bergthorsson, J.T., Ingvarsson, S., Egilsson, V. and Barkardottir, R.B. (1996) Frequent occurrence of BRCA2 linkage in Icelandic breast cancer families and segregation of a common BRCA2 haplotype. Am J Hum Genet 58, 749-56.

Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19, 1720-30.

Hadj-Slimane, R., Lepelletier, Y., Lopez, N., Garbay, C. and Raynaud, F. (2007) Short interfering RNA (siRNA), a novel therapeutic tool acting on angiogenesis. Biochimie 89, 1234-44.

Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. Cell 100, 57-70.

Hayashi, N. and Cunha, G.R. (1991) Mesenchyme-induced changes in the neoplastic characteristics of the Dunning prostatic adenocarcinoma. Cancer Res 51, 4924-30.

Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. Nat Rev Drug Discov 1, 727-30.

Hsu, H., Xiong, J. and Goeddel, D.V. (1995) The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. Cell 81, 495-504.

Hu, M. and Polyak, K. (2008) Molecular characterisation of the tumour microenvironment in breast cancer. Eur J Cancer 44, 2760-5.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M.G., Sawyer, L.R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J.P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J.F., Ellis, M.J., Olopade, O.I., Bernard, P.S. and Perou, C.M. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7, 96.

Hudis, C.A. (2007) Trastuzumab--mechanism of action and use in clinical practice. N Engl J Med 357, 39-51.

Hughes, C.M., Rozenblatt-Rosen, O., Milne, T.A., Copeland, T.D., Levine, S.S., Lee, J.C., Hayes, D.N., Shanmugam, K.S., Bhattacharjee, A., Biondi, C.A., Kay, G.F., Hayward, N.K., Hess, J.L. and Meyerson, M. (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Mol Cell 13, 587-97.

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

Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Menard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M. and Croce, C.M. (2005) MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65, 7065-70.

Jawdekar, G.W. and Henry, R.W. (2008) Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta 1779, 295-305.

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

Jonsson, G., Staaf, J., Olsson, E., Heidenblad, M., Vallon-Christersson, J., Osoegawa, K., de Jong, P., Oredsson, S., Ringner, M., Hoglund, M. and Borg, A. (2007) High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. Genes Chromosomes Cancer 46, 543-58.

Kallioniemi, A., Kallioniemi, O.P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H.S., Pinkel, D., Gray, J.W. and Waldman, F.M. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. Proc Natl Acad Sci U S A 91, 2156-60.

Kanda, T., Otter, M. and Wahl, G.M. (2001) Mitotic segregation of viral and cellular acentric extrachromosomal molecules by chromosome tethering. J Cell Sci 114, 49-58.

Kwek, S.S., Roy, R., Zhou, H., Climent, J., Martinez-Climent, J.A., Fridlyand, J. and Albertson, D.G. (2009) Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. Oncogene 28, 1892-903.

Ladomery, M. and Dellaire, G. (2002) Multifunctional zinc finger proteins in development and disease. Ann Hum Genet 66, 331-42.

Letessier, A., Sircoulomb, F., Ginestier, C., Cervera, N., Monville, F., Gelsi-Boyer, V., Esterni, B., Geneix, J., Finetti, P., Zemmour, C., Viens, P., Charafe-Jauffret, E., Jacquemier, J., Birnbaum, D. and Chaffanet, M. (2006) Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers. BMC Cancer 6, 245.

Lin, D., Sugawara, T., Strauss, J.F., 3rd, Clark, B.J., Stocco, D.M., Saenger, P., Rogol, A. and Miller, W.L. (1995) Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. Science 267, 1828-31.

Lux, M.P., Fasching, P.A. and Beckmann, M.W. (2006) Hereditary breast and ovarian cancer: review and future perspectives. J Mol Med 84, 16-28.

Ma, L., Teruya-Feldstein, J. and Weinberg, R.A. (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 449, 682-8.

Massarweh, S. and Schiff, R. (2006) Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. Endocr Relat Cancer 13 Suppl 1, S15-24.

Mattick, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., Scott, G.K., Fedele, V., Ginzinger, D., Getts, R. and Haqq, C. (2006) Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. Mol Cancer 5, 24.

McClintock, B. (1951) Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol 16, 13-47.

Miyoshi, Y., Murase, K. and Oh, K. (2008) Basal-like subtype and BRCA1 dysfunction in breast cancers. Int J Clin Oncol 13, 395-400.

Mullen, T.E. and Marzluff, W.F. (2008) 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 22, 50-65.

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

Nakamura, M., Choe, S.K., Runko, A.P., Gardner, P.D. and Sagerstrom, C.G. (2008) Nlz1/Znf703 acts as a repressor of transcription. BMC Dev Biol 8, 108.

Oh, Y.K. and Park, T.G. (2009) siRNA delivery systems for cancer treatment. Adv Drug Deliv Rev 61, 850-62.

Olayioye, M.A. (2001) Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. Breast Cancer Res 3, 385-9.

Olsen, K.E., Knudsen, H., Rasmussen, B.B., Balslev, E., Knoop, A., Ejlertsen, B., Nielsen, K.V., Schonau, A. and Overgaard, J. (2004) Amplification of HER2 and TOP2A and deletion of TOP2A genes in breast cancer investigated by new FISH probes. Acta Oncol 43, 35-42.

Osorio, A., de la Hoya, M., Rodriguez-Lopez, R., Martinez-Ramirez, A., Cazorla, A., Granizo, J.J., Esteller, M., Rivas, C., Caldes, T. and Benitez, J. (2002) Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. Int J Cancer 99, 305-9.

Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global cancer statistics, 2002. CA Cancer J Clin 55, 74-108.

Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M.J. and Petersen, O.W. (1999) Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. Dev Biol 206, 88-99.

Penault-Llorca, F., Bilous, M., Dowsett, M., Hanna, W., Osamura, R.Y., Ruschoff, J. and van de Vijver, M. (2009) Emerging technologies for assessing HER2 amplification. Am J Clin Pathol 132, 539-48.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O. and Botstein, D. (2000) Molecular portraits of human breast tumours. Nature 406, 747-52.

Peto, J. and Mack, T.M. (2000) High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 26, 411-4.

Pisani,P. http://www.euphix.org/object_document/o5614n27171.html. In.

Pole, J.C., Courtay-Cahen, C., Garcia, M.J., Blood, K.A., Cooke, S.L., Alsop, A.E., Tse, D.M., Caldas, C. and Edwards, P.A. (2006) High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. Oncogene 25, 5693-706.

Prentice, L.M., Shadeo, A., Lestou, V.S., Miller, M.A., deLeeuw, R.J., Makretsov, N., Turbin, D., Brown, L.A., Macpherson, N., Yorida, E., Cheang, M.C., Bentley, J., Chia, S., Nielsen, T.O., Gilks, C.B., Lam, W. and Huntsman, D.G. (2005) NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis. Oncogene 24, 7281-9.

Ray, M.E., Yang, Z.Q., Albertson, D., Kleer, C.G., Washburn, J.G., Macoska, J.A. and Ethier, S.P. (2004) Genomic and expression analysis of the 8p11-12 amplicon in human breast cancer cell lines. Cancer Res 64, 40-7.

Reis-Filho, J.S., Savage, K., Lambros, M.B., James, M., Steele, D., Jones, R.L. and Dowsett, M. (2006a) Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridisation analysis. Mod Pathol 19, 999-1009.

Reis-Filho, J.S., Simpson, P.T., Turner, N.C., Lambros, M.B., Jones, C., Mackay, A., Grigoriadis, A., Sarrio, D., Savage, K., Dexter, T., Iravani, M., Fenwick, K., Weber, B., Hardisson, D., Schmitt, F.C., Palacios, J., Lakhani, S.R. and Ashworth, A. (2006b) FGFR1

emerges as a potential therapeutic target for lobular breast carcinomas. Clin Cancer Res 12, 6652-62.

Rowell, N.P. (2009) Radiotherapy to the chest wall following mastectomy for node-negative breast cancer: a systematic review. Radiother Oncol 91, 23-32.

Saal, L.H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A. and Peterson, C. (2002) BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol 3, SOFTWARE0003.

Santarosa, M. and Ashworth, A. (2004) Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 1654, 105-22.

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

Shenouda, S.K. and Alahari, S.K. (2009) MicroRNA function in cancer: oncogene or a tumor suppressor? Cancer Metastasis Rev 28, 369-78.

Shimoi, W., Ezawa, I., Nakamoto, K., Uesaki, S., Gabreski, G., Aridor, M., Yamamoto, A., Nagahama, M., Tagaya, M. and Tani, K. (2005) p125 is localized in endoplasmic reticulum exit sites and involved in their organization. J Biol Chem 280, 10141-8.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A. and McGuire, W.L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235, 177-82.

Smith, I., Procter, M., Gelber, R.D., Guillaume, S., Feyereislova, A., Dowsett, M., Goldhirsch, A., Untch, M., Mariani, G., Baselga, J., Kaufmann, M., Cameron, D., Bell, R., Bergh, J., Coleman, R., Wardley, A., Harbeck, N., Lopez, R.I., Mallmann, P., Gelmon, K., Wilcken, N., Wist, E., Sanchez Rovira, P. and Piccart-Gebhart, M.J. (2007) 2-year follow-up of trastuzumab after adjuvant

chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 369, 29-36.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A.L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98, 10869-74.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lonning, P.E., Brown, P.O., Borresen-Dale, A.L. and Botstein, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100, 8418-23.

Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L. and Liu, E.T. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100, 10393-8.

Staaf, J., Jonsson, G., Ringner, M. and Vallon-Christersson, J. (2007) Normalization of array-CGH data: influence of copy number imbalances. BMC Genomics 8, 382.

Staaf, J., Ringner, M., Vallon-Christersson, J., Jonsson, G., Bendahl, P.O., Holm, K., Arason, A., Gunnarsson, H., Hegardt, C., Agnarsson, B.A., Luts, L., Grabau, D., Ferno, M., Malmstrom, P.O., Johannsson, O.T., Loman, N., Barkardottir, R.B. and Borg, A. (2010) Identification of Subtypes in Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer Reveals a Gene Signature Prognostic of Outcome. J Clin Oncol.

Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D. and Abelson, J. (2001) Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. Rna 7, 1543-53.

Stocco, D.M. and Clark, B.J. (1996) Regulation of the acute production of steroids in steroidogenic cells. Endocr Rev 17, 221-44.

Stoecklin, G., Mayo, T. and Anderson, P. (2006) ARE-mRNA degradation requires the 5'-3' decay pathway. EMBO Rep 7, 72-7.

Stratton, M.R. and Rahman, N. (2008) The emerging landscape of breast cancer susceptibility. Nat Genet 40, 17-22.

Streicher, K.L., Yang, Z.Q., Draghici, S. and Ethier, S.P. (2007) Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon. Oncogene 26, 2104-14.

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

Tang, B.L. and Ng, E.L. (2009) Rabs and cancer cell motility. Cell Motil Cytoskeleton 66, 365-70.

Tartaglia, L.A., Ayres, T.M., Wong, G.H. and Goeddel, D.V. (1993) A novel domain within the 55 kd TNF receptor signals cell death. Cell 74, 845-53.

Tenney, K. and Shilatifard, A. (2005) A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: linking leukemogensis to covalent modifications of chromatin. J Cell Biochem 95, 429-36.

Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D. and Parker, R. (2000) Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404, 515-8.

Theillet, C., Adelaide, J., Louason, G., Bonnet-Dorion, F., Jacquemier, J., Adnane, J., Longy, M., Katsaros, D., Sismondi, P., Gaudray, P. and et al. (1993) FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosomes Cancer 7, 219-26.

Thorlacius, S., Olafsdottir, G., Tryggvadottir, L., Neuhausen, S., Jonasson, J.G., Tavtigian, S.V., Tulinius, H., Ogmundsdottir, H.M. and Eyfjord, J.E. (1996) A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. Nat Genet 13, 117-9.

Toledo, F., Buttin, G. and Debatisse, M. (1993) The origin of chromosome rearrangements at early stages of AMPD2 gene amplification in Chinese hamster cells. Curr Biol 3, 255-64.

Ugolini, F., Adelaide, J., Charafe-Jauffret, E., Nguyen, C., Jacquemier, J., Jordan, B., Birnbaum, D. and Pebusque, M.J. (1999) Differential expression assay of chromosome arm 8p genes identifies Frizzled-related (FRP1/FRZB) and Fibroblast Growth Factor Receptor 1 (FGFR1) as candidate breast cancer genes. Oncogene 18, 1903-10.

Vembar, S.S. and Brodsky, J.L. (2008) One step at a time: endoplasmic reticulum-associated degradation. Nat Rev Mol Cell Biol 9, 944-57.

Verheul, H.A., Coelingh-Bennink, H.J., Kenemans, P., Atsma, W.J., Burger, C.W., Eden, J.A., Hammar, M., Marsden, J. and Purdie, D.W. (2000) Effects of estrogens and hormone replacement therapy on breast cancer risk and on efficacy of breast cancer therapies. Maturitas 36, 1-17.

Wang, Y., Pearce, M.M., Sliter, D.A., Olzmann, J.A., Christianson, J.C., Kopito, R.R., Boeckmann, S., Gagen, C., Leichner, G.S., Roitelman, J. and Wojcikiewicz, R.J. (2009) SPFH1 and SPFH2 mediate the ubiquitination and degradation of inositol 1,4,5-trisphosphate receptors in muscarinic receptor-expressing HeLa cells. Biochim Biophys Acta 1793, 1710-8.

Wojcikiewicz, R.J., Pearce, M.M., Sliter, D.A. and Wang, Y. (2009) When worlds collide: IP(3) receptors and the ERAD pathway. Cell Calcium 46, 147-53.

Yang, J., Ren, Y., Wang, L., Li, B., Chen, Y., Zhao, W., Xu, W., Li, T. and Dai, F. (2010) PTEN mutation spectrum in breast cancers and breast hyperplasia. J Cancer Res Clin Oncol.

Yang, Z.Q., Albertson, D. and Ethier, S.P. (2004) Genomic organization of the 8p11-p12 amplicon in three breast cancer cell lines. Cancer Genet Cytogenet 155, 57-62.

Yang, Z.Q., Streicher, K.L., Ray, M.E., Abrams, J. and Ethier, S.P. (2006) Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer Res 66, 11632-43.

Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F.K. and Kambach, C. (2005) Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. J Biol Chem 280, 16066-75.

Zhang, J., Liu, X., Datta, A., Govindarajan, K., Tam, W.L., Han, J., George, J., Wong, C., Ramnarayanan, K., Phua, T.Y., Leong, W.Y., Chan, Y.S., Palanisamy, N., Liu, E.T., Karuturi, K.M., Lim, B. and Miller, L.D. (2009) RCP is a human breast cancer-promoting gene with Ras-activating function. J Clin Invest 119, 2171-83.