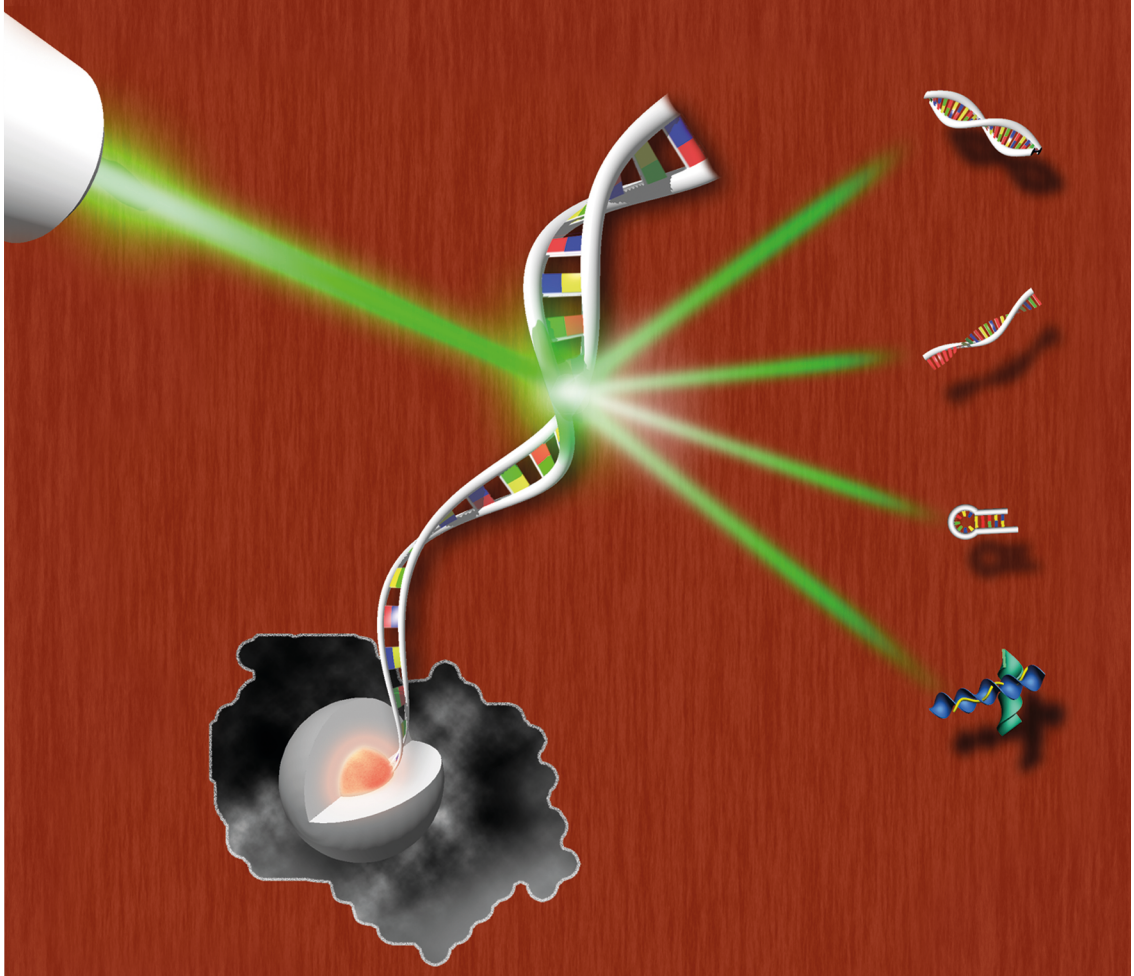
Human Breast Cancer: its Genetics, Biology and Prognosis



Muhammad Riaz

Human Breast Cancer: its Genetics, Biology and Prognosis

Muhammad Riaz

The research for this thesis was performed within the framework of the Erasmus Postgraduate School Molecular Medicine at the Department of Medical Oncology, Erasmus Medical Center, Rotterdam, The Netherlands.



Molecular Medicine
Postgraduate School

The support for the printing of this thesis was obtained by kind contribution from:
Department of Medical Oncology, Erasmus Medical Center
Erasmus University Rotterdam

Cover design and lay-out by Jean-Carlos A. Helmijr and Muhammed Riaz
Printed and bound by Ridderprint, Ridderkerk

ISBN 978-90-5335-638-8
© 2013 Muhammad Riaz

All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the author.

Human Breast Cancer: its Genetics, Biology and Prognosis

Genetische, biologische en prognostische factoren bij humane borstkanker

Thesis

to obtain the degree of Doctor from the Erasmus University Rotterdam
by the command of the rector magnificus

Prof. dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board
The public defense shall be held on
Wednesday 30th of January 2013 at 11:30 hours

By

Muhammad Riaz
Born in Pakistan



DOCTORAL COMMITTEE

Promotors:

Prof. dr. J. A. Foekens
Prof. dr. J. G. M. Klijn

Copromotor:

Dr. J. W. M. Martens

Other members:

Prof. dr. A. G. Uitterlinden
Prof. dr. Ir. M. de Jong
dr. P. M. J. J. Berns

DEDICATION

*This thesis is dedicated to my colleague late **Tjong Hauw Han (Buddy)**. May his soul rest in peace.*

CONTENTS

Chapter 1	**General Introduction**	**9**
	1.1. Epidemiology of breast cancer	**10**
	1.2. Morphology of the normal and malignant breast	**10**
	1.3. Etiology of breast cancer	**12**
	1.4. Breast cancer staging and grading	**13**
	1.5. Genetics of breast cancer	**15**
	1.6. Molecular classification of breast cancer	**23**
	1.7. Prognosis of breast cancer	**26**
	1.8. Biology of breast cancer	**27**
	1.9. Breast cancer research models	**28**
	1.10. Summary	**35**
Chapter 2	**Aims and outlines of the thesis**	**47**
Chapter 3	**Association of rare *MSH6* variants with familial breast cancer** *Breast Cancer Res Treat. 2010; 123(2):315-20*	**51**
Chapter 4	**Low-risk susceptibility alleles in 40 human breast cancer cell lines** *BMC Cancer 2009; 9:236*	**61**
Chapter 5	**Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and expression of the nearest genes** *Breast Cancer Res Treat. 2012;133(3):843-51*	**75**
Chapter 6	**High *TWIST1* mRNA expression is associated with poor prognosis in lymph node-negative and estrogen receptor-positive human breast cancer and co-expressed with stromal and ECM related genes** *Breast Cancer Res. 2012;14(5):R123*	**89**
Chapter 7	**MicroRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs** *Submitted for publication*	**111**
Chapter 8	**Growth and metastatic potential of human breast cancer cell lines in mice** *In preparation*	**133**
Chapter 9	**General discussion**	**149**
Chapter 10	**Summary/Samenvatting/ خلاصه**	**165**
Appendices	Acknowledgements List of publications PhD portfolio Author's curriculum vitae	**177**

Chapter 1

General Introduction

1.1 EPIDEMIOLOGY OF BREAST CANCER

Cancer is a major public health problem, being the second leading cause of death, after cardiovascular diseases[1]. Among women, breast cancer is the first neoplasm for incidence and the second for mortality all over the world. World-wide, an incidence of 1.4 million new cases and a mortality of 0.7 million (in both males and females) are expected in year 2012[1]. Breast cancer incidence and mortality rates strongly vary all around the world[2]. A woman's lifetime risk of developing the disease is 1 in 8 in the United States as well as in the Netherlands[3]. More than two-thirds of cases of breast cancer in 2010 were in women aged 50 years and older, most of which were in developed countries. For women aged 15–49 years, twice as many breast cancer cases were registered in developing countries than in developed countries. An increase in the number of cases in all regions during the past three decades has been registered, but this increase is more prominent in the Middle East, South Asia, South-East Asia, and Central Latin America. This could be partly due to the transition process of these countries toward a western life-style and/or improvement in diagnosis. In North America, Western Europe, and Southern Latin America the increase in breast cancer is lower than the global average. On average, the number of new cases of breast cancer increases, but a part of the recorded increase is because of the global increase of the population.

In the Netherlands, the cumulative probability of breast cancer incidence for individuals aged 15–79 years was 9.3 in 1980, 10.0 in 1990, 11.5 in 2000, and 10.5 in 2010[2]. The gradual increase is due to an increase in the average life span of the population, leaving a larger percentage of women in an age group where breast cancer is frequently diagnosed. Another reason is the introduction of the nation-wide breast cancer screening program in 1989 for women aged 50-69 years. As a result, there was a substantial increase in the incidence rates of breast cancer among women of age 50-69 years after 1989, but not for women younger than 50 years or aged 70 years and older. Additionally, among women aged 50-69 years, there was an increase in the rate of stage I cancers and a decrease in stage III+ cancer and therefore as from 1989 women aged 70-74 years were also invited for the national screening program. In contrast to this, the cumulative probability of breast cancer death for individuals aged 15–79 years decreased in the Netherlands being 3.9 in 1980, 3.9 in 1990, 3.7 in 2000, and 2.9 in 2010[2], which can be attributed to earlier diagnosis through increased awareness and the breast cancer screening program and to improved treatment[4].

1.2 MORPHOLOGY OF THE NORMAL AND MALIGNANT BREAST

1.2.1. Anatomy of the normal breast

The breast of a normal adult female is a mass of glandular, fatty and fibrous tissues located over the pactoral muscle of the chest wall attached to the chest through a strand called Cooper's ligament and occupying the entire area between the 2nd and the 6th rib of the rib cage (Figure 1.1). The terminal part of each breast has pink or brown pigmented circular area called areola surrounding a small protrusion called nipple through which milk can flow out to nurture the baby during the

entire period of lactation. The nipples are embedded with accessory mammary glands, sweat glands, and sebaceous glands that form tubercles during pregnancy that produce lubricating hormone and thus facilitate nourishment of the baby. Each breast contains 15 to 20 lobes. Each lobe is further composed of many smaller lobules, which are hormonal responsive and connected to a large number of tiny bulbs through a small duct. These bulbs are called terminal ductal lobular units (TDLUs). The TDLUs are functional parts of the mammary glands, which produce milk. The milk produced in the TDLUs is collected in the intra-lobular spaces and from where it is transported via the ductal system to the nipple. The ductal system is a branching system, which is comprised of on average 9-10 ducts that open into the nipple. Each breast is also supplied with blood and lymph vessels to feed nutrition and oxygen to the actively dividing cells and to drain out the waste material from mammary tissue. The lymph vessels lead to a series of small kidney-shaped organs called lymph nodes, which are scattered in all parts of the body including armpits, neck, and chest. Lymph nodes are also part of the breast drainage system.

1.2.2. Normal histology

The epithelium lining the entire lobular-ductal system is composed of an inner secretory luminal epithelial layer and an outer contractile basal myoepithelial layer. The inner epithelial layer consists of columnar or cuboidal shaped cells, which form a polarized continuous layer lining the entire lumen. These columnar cells are responsible for milk production and secretion into the lumen of the lobule. The outer myoepithelial layer comprises of cells, which are typically elongated and reside between luminal epithelial cells and basement membrane. These myoepithelial cells form nearly a continuous layer in the ducts, whereas they form a discontinuous basket-like structure around the lobules or acini in the lobules[5]. The myoepithelial cells also contain oxytocin receptors that are activated by the release of the hormone oxytocin from the posterior lobe of the pituitary gland[6]. The activated myoepithelial cells stimulate the contraction of muscle cells, thereby pushing secreted milk towards the nipple for the nourishment of the baby[7].

Luminal epithelial and basal myoepithelial cells are structurally distinct and can be distinguished on the basis of expression of various cytokeratin (CK) proteins and smooth muscle actin (SMA). Luminal cells express cytokeratins CK7, CK8, CK18, and CK19 and basal myoepithelial cells express CK5, CK14, CK17[8,9]. Expression of luminal CK is restricted to only luminal cells whereas expression of myoepithelial CK is restricted to only myoepithelial cells but not to the luminal cells. Additionally, myoepithelial cells have high concentrations of actin microfilaments and stain positive for SMA, whereas these actins are low or not expressed in luminal epithelial cells[7,10]. It is thought that these two distinct epithelial cell lineages of the entire ducto-lobular system are descendent of a single mammary stem cell[10-12]. In addition, two types of luminal-restricted and one myoepithelial-restricted progenitor have been identified[13,14]. However, studies are underway to elucidate the exact cellular hierarchy present in the breast epithelium.

1.2.3. Breast malignancy initiation

About 50-80% of breast cancers are presumably thought to arise in the tiny ducts of TDLUs and commonly known as ductal carcinoma whereas the remaining are lobular carcinoma, which are thought to originate from epithelial cells of lobules[15]. Breast cancer can also arise from connective and adipose tissue (sarcomas and adiposarcomas, respectively). However, these types of breast cancers are rare accounting less than 1% of the total breast cancers and therefore are not considered in this thesis. The ductal and lobular carcinomas are generally regarded as source of invasive breast cancer spreading locally and to the distant organs (Figure 1.1).

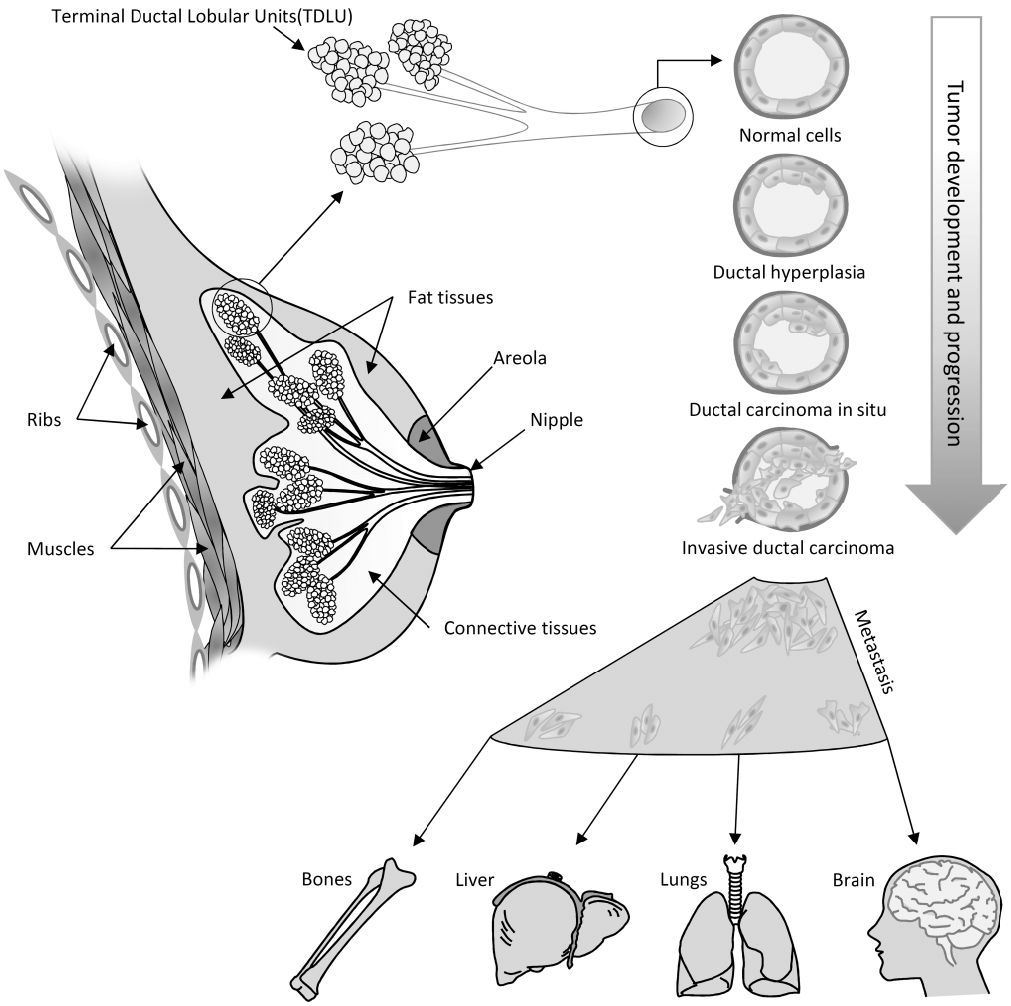


Figure 1.1 A cross section of human female breast and schematic representation of different malignant stages of breast cancer (from ductal hyperplasia to metastatic cancer).

1.3 ETIOLOGY OF BREAST CANCER

The major risk factors for breast cancer include: female gender, increasing age, western lifestyle, positive family history of breast or ovarian cancer, germline mutation in a high-risk susceptibility gene, prior diagnosis of breast cancer, benign breast disease with atypical hyperplasia, and exposure to ionizing radiation at young age. Other risk factors are related to endogenous and/or exogenous hormone exposure due to early age at menarche, late menopause, older age at first

childbirth, nulliparity, small number of children, use of oral contraceptives, high socioeconomic status, dense breast tissue on mammography and high postmenopausal obesity body mass index, and waist-hip ratio[16-19]. The hormone exposure-related factors showed expected patterns of association with estrogen receptor (ER) positive (+)/progesterone receptor (PGR) positive (+) but not with ER-negative (-)/PGR+ or ER-/PGR- breast cancers[20]. Genetic factors and their role, especially in the development of hereditary breast cancers, are extensively discussed in the section 1.6 "Genetics of breast cancer" of this chapter.

1.4 BREAST CANCER STAGING AND GRADING

1.4.1. Tumor staging

Tumor staging is a standard way of summarizing the size of the tumor and how far cancer has spread in the body. The most common system used to describe the breast cancer stages is the TNM system developed by the American Joint Committee on Cancer (AJCC)[21]. In the TNM system T refers to tumor size, N indicates whether or not tumor is spread to the lymph nodes, and M indicates whether tumor has spread to distant organs. Based on these three features, breast tumors are classified into five stages from 0 to IV. Stage 0 refers to a carcinoma in situ with no regional lymph node metastasis and it has 93% 5-year survival rate whereas stage IV describes a disease in which primary tumor has spread to the lymph nodes and a detectable metastases and distant organ is evident. Five year survival rate of this stage is poor and generally less than 15%. Tumor stages are strong predictors of prognosis of breast cancer. Tumor stages based on TNM classification along with brief description is given in Table 1.1 below.

1.4.2. Tumor grading

The grading of breast tumor is a term used to describe how closely a tumor resembles the normal tissue, which is an indication how quickly a cancer can progress. There are several grading systems available to assign a grade to breast tumors, however, the Scarff-Bloom-Richardson grading system, modified by Elston and Ellis is widely accepted and frequently used all over Europe[22]. According to this system, a tumor is inspected by a trained pathologist and a grade is given based on three microscopic features of the tumor: (i) degree of tubule or gland formation, (ii) nuclear pleomorphism, and (iii) mitotic count (Table 1.2). Each feature is scored on a scale from 1-3. After combining these feature scores, breast tumor can be of three grades: well differentiated (grade I), moderately differentiated (grade II), and poorly differentiated (grade III)[23]. Tumor grade has been demonstrated as an independent prognostic factor by numerous studies and high-grade tumors tend to show poor prognosis compared with grade II tumors, which demonstrate intermediate prognosis.

Table1.1 Staging of human breast cancer[21]

Stage	TNM system	Description	5-year survival
0	Tis N0 M0	Carcinoma in situ, no regional lymph node metastasis, no distant organ metastasis.	93%
I	T1 N0 M0	Tumor is $\leq$ 2 cm in greatest dimension, no regional lymph node metastasis and no distant organ metastasis.	88%
IIA	T0 N1 M0	No evidence of primary tumor, metastasis to movable ipsilateral nodes, no distant organ metastasis.	81%
	T1 N1 M0	Tumor is less than or equal to 2 cm, metastasis to movable ipsilateral node, no distant organ metastasis.	
	T2 N0 M0	Tumor is between 2 and 5 cm, no regional lymph node metastasis and no distant organ metastasis.	
IIB	T2 N1 M0	Tumor is between 2 and 5 cm, metastasis to movable ipsilateral nodes, no distant organ metastasis.	74%
	T3 N0 M0	Tumor is > 5 cm, no tumor in regional lymph node, no distant organ metastasis.	
IIIA	T0 N2 M0	No evidence of primary tumor, metastasis to fixed ipsilateral nodes, no distant organs metastases.	67%
	T1 N2 M0	Tumor is $\leq$ 2 cm, metastases to fixed ipsilateral nodes, no distant organs metastases.	
	T2 N2 M0	Tumor is between 2 and 5 cm, metastases to fixed ipsilateral nodes, no distant organs metastases.	
	T3 N2 M0	Tumor is > 5 cm, metastasis to movable or fixed ipsilateral nodes, no distant organ metastasis.	
IIIB	T4 N0 M0	Tumor of any size with direct extension to the chest wall or skin tissue, no evidence of regional lymph node as well as distant organ metastasis.	41%
	T4 N1 M0	Tumor of any size with direct extension to the chest wall or skin tissue, no metastasis to movable ipsilateral nodes, no distant organ metastasis.	
	T4 N2 M0	Tumor of any size with direct extension to the chest wall or skin tissue, metastasis to fixed ipsilateral nodes, no distant organ metastasis.	
IIIC	AnyT N3 M0	Any primary tumor involvement, metastasis in ipsilateral intraclavicular lymph nodes, no distant organ metastasis.	49%
IV	Any T Any N M1	Any primary tumor involvement, any nodal involvement, distant detectable metastasis as determined by classical and radiographic means.	<15%

T = Status of primary tumor; N = Regional lymph nodes; M = Distant metastasis.

Table 1.2 Grading of breast cancer[23]

Tubule or Gland formation (% of carcinoma composed of tubular structures)	**Score**
More than 75% of the tumor contains tubule formation	1
10 to 75% of the tumor has tubule formation	2
Less than 10% of the tumor contains tubule formation	3

Nuclear pleomorphism (change in cells)	**Score**
Nuclei are small and uniform in shape and size	1
Nuclei are moderate in size and variation	2
Nuclei possess marked variation in size and shape with multiple prominent nucleoli	3

Mitosis count (per 10 high-power fields (hpf) with field area of 0.274 mm^2)	**Score**
0 to 9 cell divisions	1
10-19 cell divisions	2
Equal or more than 20 cell divisions	3

Tumor grade	**Brief description**	**Feature scores**
I	Well-differentiated (normal appearing, slowly growing, not aggressive).	3-5
II	Moderately-differentiated (semi-normal, growing moderately fast, and intermediately aggressive).	6-7
III	Poorly differentiated (abnormal, very fast growing, and very aggressive).	8-9

1.5 GENETICS OF BREAST CANCER

Breast cancer is a genetic disease that arises due to abnormalities (mutations) in one or more rate-limiting genes in cells. There are two basic types of genetic mutations: germline and acquired. If the mutation is passed directly from a parent to a child, it is called a germline mutation. This would mean that the mutation is present in all cells of an individual including the reproductive cells. Breast cancer caused by germline mutations is called inherited breast cancer (hereditary breast cancer), which is less common and it accounts for about 20% of all breast cancer cases. On the other hand, the acquired mutations occur in a single cell of the target tissue during a person's life, and they are passed on to all cells that originate from that single affected cell. These mutations are also known as somatic mutations and they are generally caused by various environmental factors, such as UV radiations, viruses, cigarette smoke, and alcohol consumption. Breast cancer caused by somatic mutations is called sporadic cancer. Sporadic breast cancer is more frequent and it accounts for the remaining 80% of breast cancer cases. Mutated genes in breast cancer are generally classified into two groups depending on their effect or gain and loss functions. One group includes genes known as tumor suppressor genes, which, when mutated, are unable to control unnecessary growth of cells and thereby cause cancer. It is important to note that the tumor suppressor genes are generally recessive and therefore their complete loss of function requires both alleles to be affected. The other and opposite group includes genes known as

oncogenes, which, when mutated, produce an active variant of the proteins that constitutively induce cell proliferation or inhibition of cell death and thereby leading to cancer. The oncogenes are dominant and activations of them generally require only one allele to be mutated.

1.5.1. Hereditary breast cancer

A family history of breast cancer is present in about 20% of women with newly diagnosed disease. The suspicion of hereditary breast cancer is increased if a family tree includes multiple cases of breast cancer, breast cancer occurring below the age of 50 year, cases with bilateral breast cancer, male relative with breast cancer, and cases with cancer in the ovary, fallopian tube, or primary peritoneum. Also, individuals of Ashkenazi Jewish heritage are at increased risk of having hereditary breast cancer, particularly a disease with a *BRCA1* mutant gene. Family members of such individuals can be considered at an increased risk of developing breast cancer and may be considered for more intense screening or genetic counseling. In the last decade of the 20th century, several high-risk breast cancer susceptibility genes have been discovered[24-26]. More recently, with the advent of array technologies and using case control studies numerous moderate- and low-risk breast cancer susceptibility loci have been discovered[27].

1.5.1.1. High-risk breast cancer susceptibility genes

Linkage studies in families have identified mutations in genes, which are associated with high-risk of breast cancer. These high-risk breast cancer susceptibility genes include *BRCA1, BRCA2, TP53, PTEN, STK11,* and *CDH1* (Table 1.3).

Table 1.3 High-risk breast cancer susceptibility genes

Locus	Genes in/near region	Variants*	MAF	RR	References
17q21.0	*BRCA1*	c.185delAG, c.5382insC	0.0006	5-45	[31], [32], [33]
13q12.3	*BRCA2*	c.6174delT	0.001	9-21	[32], [34]
17p13.1	*TP53*	c.13964G>C	rare	2-10	[35]
10q23.3	*PTEN*	p.L22X, p.Q109X, p.R188X, p.V332E, p.R334X, p.R232X	rare	2-10	[36], [37]
19p13.3	*STK11*	p.Y60X, p.C759A, p.G2412A, c.716delGGTC, c.843delG	rare	2-10	[38], [30]
16q22.1	*CDH1*	p.W20X, p.Q23X, c.1009-1delAG, c.1075insA	rare	2-10	[29], [39]

*The most commonly identified variants in breast cancer families. MAF: Minor Allele Frequency. RR: Relative Risk.

Oncogenic mutations in one of these genes predispose women to 60-80% life time risk of developing breast cancer compared to approximately 10% breast cancer risk estimated for the general population. *BRCA1* and *BRCA2* are tumor suppressor genes, which predispose to both breast and ovarian cancers. Germline mutations in these two genes are very rare in the general population with carrier frequency of about 0.1%, yet these are identified in about 20% breast cancer families[24-26]. Besides BRCA1/2 genes, germline mutations in other high-risk genes are associated with syndromes. For example, mutations in *TP53* cause Li Fraumani syndrome, in *PTEN*

cause Cowden syndrome, in *STK11* cause Peut Jegher syndrome, and in *CDH1* cause gastric syndrome. Mutations in these genes altogether account for less than 1% of all familial breast cancer cases[24,28-30].

1.5.1.2. Moderate-risk breast cancer susceptibility genes

Moderate-risk breast cancer susceptibility genes have often been identified by a combination of candidate-based and population-based approaches. Using these approaches, six moderate-risk breast cancer susceptibility genes have been identified. These genes include *CHEK2, ATM*, *BRIP1*, *PALB2, RAD50*, and *NBS1*[40-45] (Table 1.4). The *CHEK2* gene is activated in response to DNA damage[46]. Germline variants in *CHEK2* gene including 1100delC allele have been associated with a 2 to 4-fold increased risk of breast cancer. Germline variants in ATM gene cause ataxia-telangiectasia and have been associated with a 2-fold increased risk of breast cancer risks. The *BRIP1* and *PALB2* genes encode BRCA1-interacting and BRCA2 interacting proteins, respectively. Studies have found many germline variants in these two genes to be associated with about a 2-fold higher risk of breast cancer risks. Both the *RAD50* and *NBS1* genes encode proteins that make MRE11–RAD50–NBS1 protein complexes, which is critical for proper maintenance of genomic integrity and tumor suppression. Germline variants in these genes confer to 2 to 3-fold increased risk of breast cancer. As a whole, women inheriting mutations in one of these moderate-risk breast cancer susceptibility genes predispose 20-40% elevated risks of developing breast cancer compared with the 10% in the general population. Altogether, these genes explain about 5% of familial breast cancer and 1 % of total breast cancer cases.

Table 1.4 Moderate-risk breast cancer susceptibility genes

Locus	Genes in/near region	Variants*	MAF	RR	References
11q22.3	*ATM*	c.790delT, c.170G > A, c.687delA,	0.003	2-3	43, 47, 48
22q12.10	*CHEK2*	p.I157T, c.1100delC, p.S428F,	0.004	2-3	42 , 46
17q22-24	*BRIP1*	c.141delC, p.Q944E, c.2392C > T	0.001	2-3	44, 49
16p12.1	*PALB2*	p.G796X, p.W1038X, p Y1183X	rare	2-4	45
5q31.1	*RAD50*	p.Q350X, c.687delT,	rare	2-3	40, 50
8q21.3	*NBS1*	c.657del5	rare	2-4	51

*The most commonly identified variants in breast cancer families. MAF: Minor Allele Frequency. RR: Relative Risk.

1.5.1.3. Low-risk breast cancer susceptibility loci

Low-risk breast cancer susceptibility loci are common in the general population with population frequency ranging from 5 to 50%[24]. These low-risk breast cancer susceptibility loci are scattered all over the genome and usually identified through genome-wide association studies (GWAS). GWAS have analyzed hundreds of thousands of single nucleotide polymorphisms (SNPs) as markers capturing genome-wide genetic variability of samples from affected (cases) and normal (controls) individuals[52,53]. Generally, GWAS are performed in three stages, with a relatively small initial cohort, which are then replicated in independent second and third cohorts with up to tens of thousands of

cases and controls. The identification of SNPs associated with disease in GWAS studies is based on a large number of statistical tests, which increases false discoveries and therefore stringent acceptance levels for statistical significance are used[54]. For example, a threshold of $P < 5x10^{-8}$ is considered to be an appropriate threshold for one million analyses. Another important aspect of GWAS is the availability of large cohorts of cases and controls. The requirement of large study cohorts has led to the establishment of international consortia, e.g. Breast Cancer Association Consortia (BCAC), The Consortium of Investigators of Modifiers of BRCA1/2(CIMBA), Breast Cancer Family Registry (BCFR), Baltic Familial Breast Ovarian Cancer Consortium (BFBOCC), and Breast Cancer Surveillance Consortium (BCSC) etc. For example, BCAC includes more than 68 international collaborative research groups, with a combined samples size of about 70,000 breast cancer cases and 68,000 controls[27]. To date, 21 low-risk breast cancer susceptibility loci have been convincingly identified to be associated with modest-breast cancer risks of ~ 1.2 to 1.3-fold (Table 1.5). Altogether, these loci are estimated to account for approximately 9% of familial breast cancer cases and 2% of breast cancer cases overall[27,55].

Table 1.5 Low-risk breast cancer susceptibility loci

Locus	Genes in/near region	Variants	MAF	RR	References
10q26	*FGFR2*	rs2981582	0.38	1.26	[56]
16q12	*TOX3*	rs3803662	0.25	1.20	[56]
5q11	*MAP3K1*	rs889312	0.28	1.13	[56]
8q24	*FAM84B/MYC*	rs13281615	0.40	1.08	[56]
11p15	*LSP1*	rs3817198	0.30	1.07	[56]
3p24	*NEK10/SLCA4A7*	rs4973768	0.46	1.11	[57]
17q23.2	*COX11*	rs6504950	0.27	0.95	[57]
10p14	*CASP8(D302H)*	rs1045485	0.13	0.88	[58]
2q35	*TNP/IGFBP2,5/TNS1*	rs13387042	0.52	1.12	[59]
1p11.2	*NOTCH2/FCGR1B*	rs11249433	0.40	1.14	[60]
14q24.1	*RAD51L1*	rs999737	0.24	0.84	[60]
5p12	*MRPS30/FGFR10*	rs10941679	0.26	1.19	[61]
6q25.1	*ESR1*	rs2046210	0.35	1.29	[62]
9p21.3	*CDKN2A, CDKN2B*	rs1011970	0.17	1.29	[63]
10p15	*ANKRD16, FBXO18*	rs2380205	0.43	0.89	[63]
10q21	*ZNF365*	rs10995190	0.15	0.83	[63]
10q22.3	*ZMIZ1*	rs704010	0.39	1.13	[63]
11q13	Intergenic	rs614367	0.15	1.27	[63]
12p11	*PTHLH*	rs10771399	0.12	0.79	[27]
12q24	*MK5 or TBX3*	rs1292011	0.41	0.88	[27]
21q21	*RIP140*	rs2823093	0.26	0.96	[27]

MAF: Minor Allele Frequency. RR: Relative Risk.

1.5.1.4. Low-risk breast cancer loci and polygenic susceptibility

In addition to high-risk genes like *BRCA1*and *BRCA2* and *TP53*, several low-risk breast cancer genes or loci are found to be associated with increased breast cancer risk (see section 1.5.1.3). These susceptibility alleles are common in the population and carry essentially low effect size (per-allele relative risks of 0.83–1.3)[27]. This has suggested that breast cancer disease is not predisposed by merely an individual gene, but many common genetic variants contribute to the disease susceptibility. It is therefore more likely that the inheritance of breast cancer is polygenic in which a number of susceptibility variants combined additively or multiplicatively confer increased breast cancer risk. According to this model, a woman harboring several susceptibility alleles would be at high risk compared to one with no or few breast cancer susceptibility variants. This model has been found to be well-fitting to genetic data obtained from a series of non-*BRCA1*and *BRCA2* mutant multiple-case families[64]. It is probably also an appropriate model for many other cancers and other complex diseases. The model is based on the assumption that the risk conferred by the common genetic variants in a population is predicted to be log-normally distributed. This will provide sufficient information to set up breast cancer screening programs in such a way that women harboring large numbers of risk-loci would be subjected to breast cancer screening such as mammography at early age compared to those harboring no or few risk-loci[65,66]. This would eventually allow a targeted screening and early diagnosis of breast cancer.

1.5.1.5. The biology of low-risk breast cancer susceptibility loci

The biological mechanisms by which low-risk breast cancer susceptibility loci confer their increased breast cancer risks are poorly understood. Although, the identified SNPs have been shown to be significantly associated with increased breast cancer risk, yet it is not clear whether the SNPs by themselves are causal or whether they associate with causal variants in the surrounding. To unravel exact biological disease mechanisms, various approaches have been applied. The first approach included the determination of mutant variants through deep sequencing of the entire coding genes containing or located close to these risk SNPs and subsequently correlating mutations to the corresponding risk SNP. Unfortunately, this approach has not yet yielded any causal variants in the coding part of the genome[56]. The second approach was setting-up functional studies on the identified loci. The rational of this approach was based on the observation that the majority of the low-risk loci (for example *FGFR2*, *MAP3K1*, 8q24 and 5p loci) are strongly correlated with ER-positive disease, some loci, such as *TOX3*, are associated with both ER-positive and ER-negative diseases and also some with triple-negative breast cancer[56,59,67,68]. The *FGFR2* gene is highly expressed in ER-positive compared with ER-negative cell lines and tumors and has been implicated in the genesis of ER-positive tumors[69-74]. This approach worked out successfully for only *FGFR2* and 8q24 risk loci. In case of the *FGFR2* risk variant, the risk-associated SNP rs2981582 (located in intron 2 of *FGFR2*) was itself found to be a causal variant as this variant regulates increased *FGFR2* expression by altering the binding affinity of transcription factors Oct1/Runx2 and C/EBPβ[75]. In case of the 8q24 risk variant (rs13281615), it was functionally shown that risk is conferred by expression regulation of the *MYC* gene through a tissue-specific physical interaction of enhancers from the 8q24 region[76].

The third approach was to analyze the association of low-risk loci with expression modulation of genes that either contain these loci or located nearby to them on the same linkage-disequilibrium block. This approach revealed only a significant association of the rs3803662 risk variant with lower expression of the *TOX3* gene (previously known as *TNRC9*) indicating a tumor suppressor role of the *TOX3* gene (chapter 4 of this thesis). However, functional support of this observed association is required. The *TOX3* gene encodes a high mobility group box nuclear protein involved in calcium-dependent transcription and structural modification of chromatin. Altogether, these data support the notion that the low-risk breast cancer loci do not confer their breast cancer risk through one general mechanism, but rather multiple disease mechanisms can be envisioned. However, more ongoing and future sequencing, fine-mapping and functional studies will help in elucidating the most relevant biological mechanisms underlying these common low-risk breast cancer loci.

1.5.2. Sporadic breast cancer

Approximately 80% of breast cancers are sporadic tumors (Figure 1.3). The sporadic breast cancers arise due to somatic mutations in functionally important genes. Whole genome sequencing of the consensus coding sequences of the breast cancer genome has revealed a variety of somatic mutations in large numbers of genes[77-79]. These mutations include both synonymous (not altering an amino acid) and nonsynonymous (altering an amino acid) mutations. The nonsynonymous mutations comprised of missense and nonsense mutations, insertions, deletions, rearrangements, copy number alterations and mutations in splice sites, in the regulatory elements, and in UTRs. In sporadic breast cancer, somatic mutations confer selective growth advantage to cells and therefore promote tumorigenesis. The causal somatic mutations are known as 'driver' mutations and they typically occur through misincorporation of nucleotides during DNA replication or through exposure to exogenous and endogenous mutagens. In contrast, some somatic mutations are biologically inert and do not confer any growth advantage to the cells. Such mutations are called 'passenger' mutations. These mutations normally occur in cells that have already acquired growth advantage. Mutational activation of oncogenes, coupled with mutational inactivation of tumor suppresser genes is an early event in the development of a sporadic tumor, which might be followed by independent mutations in a dozen of genes along the line of breast cancer progression.

1.5.2.1. Prevalence of driver and passenger mutations in cancers

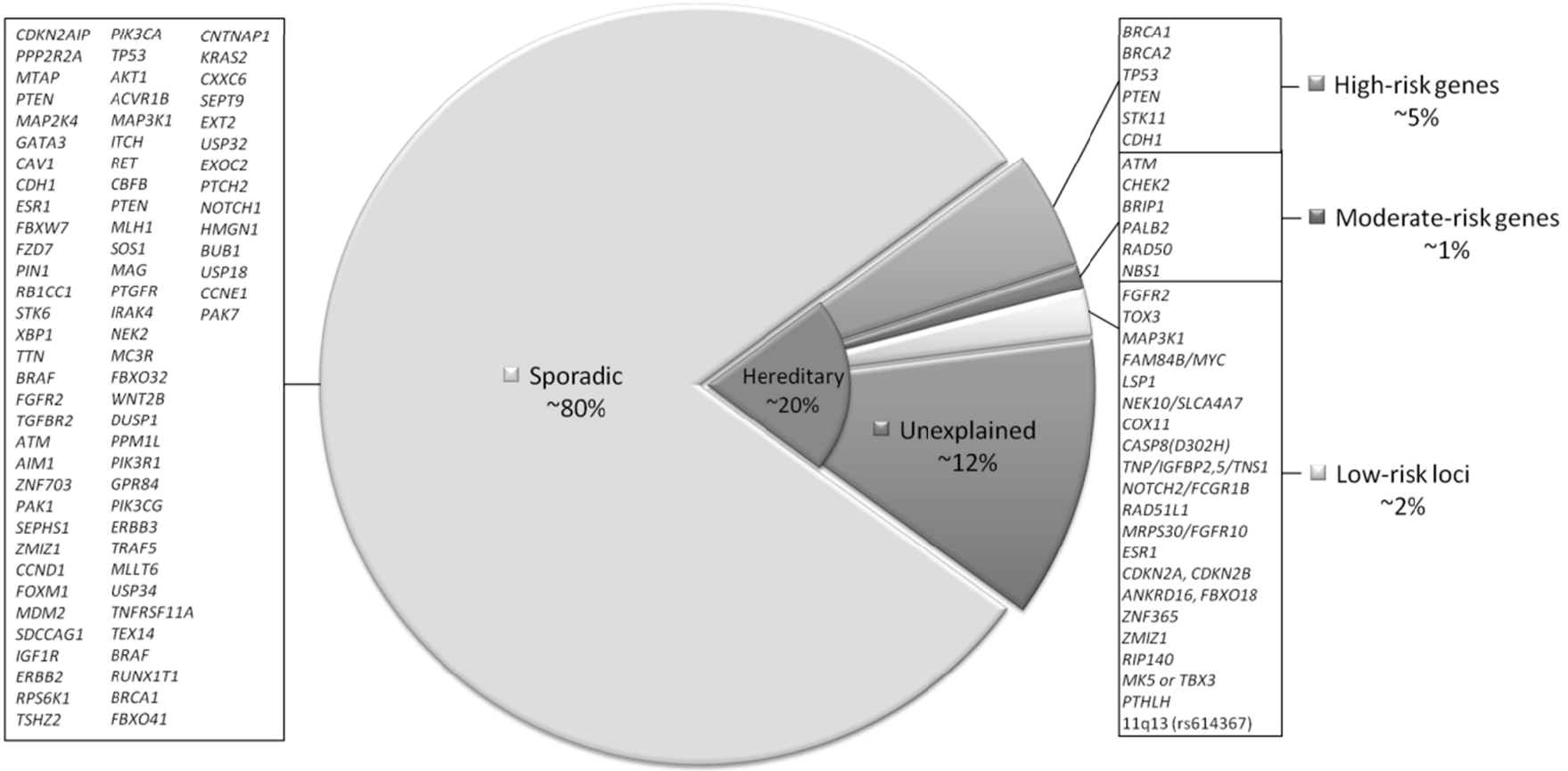
Greenman and his coworkers, for first time, investigated numbers of driver and passenger mutations by sequencing the coding exons of the 518 kinase genes in 210 diverse human cancers[77]. They compared an observed ratio of non-synonymous/synonymous mutations with that expected by chance with an underlying hypothesis that biological selection is exerted on non-synonymous mutations because these carry functional implications. Conversely, synonymous mutations are generally biologically neutral and therefore cannot be selected for biological consequences. The presence of a driver or a passenger mutation is inferred by determining a ratio of non-synonymous and synonymous mutations. A higher mutational ratio compared to that

expected by chance indicates overall positive selection (selection pressure <1), which suggests a driver mutation whereas a lower non-synonymous/synonymous ratio indicates no overall selection (selection pressure <1) and is suggestive of a passenger mutation. To distinguish genes that are statistically likely to contribute to tumorigenesis from those in which passenger mutations occur by chance, Sjoblomand and coworkers developed a statistical method that provides a cancer mutation prevalence (CaMP) score as an output for each gene analyzed[80]. According to their study, 90% of the genes with CaMP score >1.0 were predicted to have mutation frequencies higher than the background mutation frequency[80].

Using these criteria, large numbers of driver and passenger mutations have been estimated from a number of studies. Greenman and coworkers have found 921 base substitution mutations identified in 518 kinase genes, out of which 763 mutations were estimated to be passenger mutations and are not likely implicated in cancer development. The remaining 198 mutations were therefore estimated to be driver mutations, which were predicted to be distributed over 119 kinase genes[77]. These results, therefore, provide a statistical support for a large set of protein kinase genes to be implicated in one-third of the cancer cases. Curtis and co-workers have analyzed somatic copy number variations (CNVs) in 2000 breast cancers and introduced a new approach to identify new deriver genes[81]. This approach involves an identification of CNVs through an integrated genomic/transcriptomic analysis of breast cancers and then their association with the expression changes of genes located in *cis* and *trans*. This has revealed several known and unknown putative driver genes including deletions in *PPP2R2A*, *MTAP*, and *MAP2K4*[81].

1.5.2.2. Breast cancer candidate genes

About 122 breast cancer candidate genes (CAN-genes) have previously been identified and assigned to a functional group based on Gene Ontology molecular groups or biological processes groups[77]. Functional groups identified through this way are of special interest, for instance 22 of the 122 (18%) breast CAN-genes are transcriptional regulators. In the same study, it was revealed that at least one of these CAN-genes is mutated in more than 80% of the breast tumors[82]. The mutation prevalence in CAN-genes has been evaluated more precisely in a cohort of 183 breast cancer tumors (59 = ERBB2-positive, 65 = HR+ (ER/PGR-positive), and 59 = TNBC (ER/PGR/ERBB2-negative). This study identified 10, 6, and 5 genes that harbor significantly prevalent protein altering mutations in hormone receptor positive tumors, ERBB2-positive and triple-negative tumors, respectively[78]. A recent mutation analysis study has identified 32 CAN-genes in primary TNBC whose expression was significantly influenced by the presence of somatic mutations[83]. Among these genes *TP53* and *PIK3CA* are the most variable and somatically mutated genes[83]. These mutations have been evaluated for their possible effect on predicted protein function of the CAN-genes and pathways in which these CAN-genes were involved in breast cancer[84]. For example, in breast cancer, *PIK3CA* or *PTEN*, and *TP53* are frequently mutated, which obstruct PI3K and p53 signaling pathways, respectively[83]. Altogether, somatic mutations in the CAN-genes may explain the susceptibility for breast cancer associated with non-familial human breast cancer (Figure 1.2).


CDKN2AIP
PPP2R2A
MTAP
PTEN
MAP2K4
GATA3
CAV1
CDH1
ESR1
FBXW7
FZD7
PIN1
RB1CC1
STK6
XBP1
TTN
BRAF
FGFR2
TGFBR2
ATM
AIM1
ZNF703
PAK1
SEPHS1
ZMIZ1
CCND1
FOXM1
MDM2
SDCCAG1
IGF1R
ERBB2
RPS6K1
TSHZ2
PIK3CA
TP53
AKT1
ACVR1B
MAP3K1
ITCH
RET
CBFB
PTEN
MLH1
SOS1
MAG
PTGFR
IRAK4
NEK2
MC3R
FBXO32
WNT2B
DUSP1
PPM1L
PIK3R1
GPR84
PIK3CG
ERBB3
TRAF5
MLLT6
USP34
TNFRSF11A
TEX14
BRAF
RUNX1T1
BRCA1
FBXO41
CNTNAP1
KRAS2
CXXC6
SEPT9
EXT2
USP32
EXOC2
PTCH2
NOTCH1
HMGN1
BUB1
USP18
CCNE1
PAK7
Sporadic
~80%
Hereditary
~20%
Unexplained
~12%
BRCA1
BRCA2
TP53
PTEN
STK11
CDH1
High-risk genes
~5%
ATM
CHEK2
BRIP1
PALB2
RAD50
NBS1
Moderate-risk genes
~1%
FGFR2
TOX3
MAP3K1
FAM84B/MYC
LSP1
NEK10/SLCA4A7
COX11
CASP8(D302H)
TNP/IGFBP2,5/TNS1
NOTCH2/FCGR1B
RAD51L1
MRPS30/FGFR10
ESR1
CDKN2A, CDKN2B
ANKRD16, FBXO18
ZNF365
ZMIZ1
RIP140
MK5 or TBX3
PTHLH
11q13 (rs614367)
Low-risk loci
~2%

1.6 MOLECULAR CLASSIFICATION OF BREAST CANCER

1.6.1. Steroid hormone receptor expression

One of the earliest molecular classifiers of the breast tumors is the presence of the hormonal receptor molecules ER and PGR in the tumors. Approximately 75% of breast cancers are positive for ER, PGR, or both, and estrogen stimulation of ER is a significant factor in the development and growth of breast cancer. Few (<1%) of breast tumors are ER negative and PGR positive and about 15% of all breast cancers lack expression of both hormone receptors[85]. The ER and PGR expression is determined by multiple techniques among which immunochemistry (IHC) is the preferred technique and currently being used in the clinic. Generally the tumors are considered positive for ER or PGR if $\geq$10% of nuclei are stained positive immunohistochemically. The presence of ER and PGR are significantly negatively correlated with histological grade and its mitotic index. ER and PGR positive tumors are commonly grade I, well-differentiated tumors, and they are generally less aggressive compared to ER-negative tumors, which are often poorly differentiated. ER is a target for endocrine therapy such as tamoxifen, which inhibits the growth of breast cancer through competitive antagonism of estrogen at ER[86]. Tamoxifen is used for premenopausal patients with ER-positive disease. The second therapeutic agent to treat ER-positive postmenopausal breast cancers are aromatase inhibitors, which specifically blocks the aramotase-mediated conversion of androgen to estrogen and therefore less estrogen is available to stimulate ER-positive tumor cells in postmenopausal women[87].

1.6.2. ERBB2 expression

The *ERBB2* oncogene encodes a transmembrane tyrosine kinase receptor, which has evolved as a major classifier of invasive breast cancer. It is amplified and/or overexpressed in approximately 16% of breast cancer in the Netherlands. The ERBB2 overexpression is routinely determined by IHC and/or fluorescent *in-situ* hybridization (FISH) techniques in all cases of invasive breast carcinoma, although variability in testing remains a significant issue with both methodologies. Tumors are considered ERBB2 overexpressed on IHC, when at least 10% of cells show strong membranous staining and/or on FISH when a gene copy-to–CEP-17 ratio is found greater than 2. Clinically, ERBB2 overexpression has been associated with increased tumor aggressiveness, increased recurrence rates, and increased mortality rates in node-positive patients, while this relationship in node-negative patients is more variable[88-92]. Histologically, ERBB2 overexpressing tumors are often intermediate or poorly differentiated with the presence of lymph-node metastases at presentation[93]. According to recent molecular classification of breast cancer, tumors with a positive ERBB2 status do not constitute a unique molecular category and are identified in both the ER-positive and ER-negative tumor classes[94]. ERBB2-positive breast cancer is treated by monoclonal antibodies such as trastuzumab, which is directed against ERBB2 protein[95,96]. Nevertheless, some tumors develop resistance to trastuzumab and the therapy may fail and therefore more treatments options are constantly being investigated[97].

1.6.3. Molecular subtypes of breast cancer

Microarrays-based gene expression profiling of breast cancer has provided a novel working model for the taxonomy of breast cancer, which is based on gene expression analysis of breast tumors and classifying them into several molecular subtypes with different prognosis and therapeutic responses[98]. A microarray includes a collection of DNA fragments (cDNA or oligonucleotides) spotted on a solid surface (a 'chip') in an ordered pattern of grids. Each spot consists of a high concentration of unique fragments of DNA, which hybridize to a specific gene sequence present in a mixture of cDNAs derived by reverse transcription of mRNA extracted from a tumor sample [99-101]. The intensity signal of cDNA hybridized to their corresponding sequences spotted on the array is an indication of the expression level of the genes in the sample. In this way, global expression patterns of hundreds and thousands of genes in a large number of tumor sample and normal tissues are determined and consequently tumor samples are classified based on their similar gene expression patterns.

1.6.3.1. Five intrinsic molecular subtypes

For the first time, using cDNA microarrays, Perou and coworkers determined gene expression profiles of invasive breast cancers, which included paired samples from the same tumors as well[102,103]. Unsupervised cluster analysis revealed two main clusters of the tumors, which were primarily discriminated into ER-positive and ER-negative tumors. The expression patterns of both types of the tumors were remarkably different. Additionally they found a set of 496 "intrinsic genes" whose expression differed most between tumors compared to replicates from the same tumors. Intrinsic gene expression divides both ER-positive and ER-negative invasive breast tumors into five distinct molecular subtypes. These five subtypes include: luminal A and B; ERBB2-positive; basal-like; and normal-like[102-104] (figure 1.2). With the advancement of array technology and bioinformatics tools, more and more gene expression profiles have been generated and consequently new molecular subtypes of the breast cancer have been identified (as discussed below)

The tumors in luminal A and B subtypes are characterized by high expression of ER and ER-regulated genes (*GATA3*, *XBP1*, *TFF3*, *HNF3α,* and *LIV1)* and low expression of myoepithelial cell-associated genes (*KRT5*, *KRT17*, *ANXA8*, *CXCL1*, and *TRIM29)* and proliferation related-genes (*Ki-67*). The difference between the luminal A and the luminal B subtypes is that the tumors in the luminal-B subtype have comparatively low expression of ER and ER-regulated genes and high expression of the Ki-67 gene and other proliferated genes. ERBB2-positive subtype tumors are characterized by higher expression of the *ERBB2* gene and other genes transcribed from this genomic region (Entrez Gene cytogenetic band*:* 17q11.2-q12) including *GATA4*, *GRB7, NFKB* and high expression of proliferation related genes and low expression of luminal genes. The ERBB2-positive tumors are mainly hormone receptor negative but a substantial number of ERBB2-positive tumors are ER-positive and usually cluster together with the luminal B tumors. The basal-like tumors are characterized by low or no expression of luminal and *ERBB2* gene clusters (*KRT8, KRP18*, *GATA3*, *XBP1*, *TFF3*, and *HNF3α*) and high expression of the typical basal gene cluster (*KRT5*, *KRT14, KRT17*,

ITGβ4, LAMA3, *c-Kit*, *ITGα6*, *MT1X*, *CDH3*, *B-FABP*)[102,103]. Also, basal-like tumors often exhibit high expression of *EGFR*. Generally, basal-like tumors display poor prognosis and can only be treated with chemotherapeutics as no drug-related targets are known for these tumors. The normal-like subtype is the smallest and a poorly characterized group of breast cancer. The main characteristic of these tumors is that they consistently cluster together with normal breast tissue samples and fibroadenomas[102]. Importantly the intrinsic subtyping of breast tumors is very robust classification method as the existence of these five intrinsic subtypes has consistently been replicated by a number of subsequent studies using various array platforms and patient series with different ethnic background[88,105,106].

Several studies have shown that the intrinsic subtyping of the breast tumors carries prognostic significance[107-109] for example luminal A and luminal B tumors show distinct but better prognostic behavior compared to basal-like tumors and ERBB2-positive tumors. Furthermore, the basal-like subtype is enriched for tumors lacking the expression of ER, PGR, and ERBB2 expression (triple-negative). Intrinsic subtyping also has predictive power since basal-like and ERBB2-positive tumors respond more frequently to neo-adjuvant chemotherapy than other subtypes but they are more refractory to endocrine agents. Moreover, the intrinsic subtypes of breast cancer are also associated with the ability to disseminate to different distant organs[110,111]. Breast tumors of the luminal intrinsic subtypes more frequently relapse to bone and pleura, whereas tumors of the basal-like and ERBB2-positive subtypes frequently metastasize to the brain and liver[110,112].

1.6.3.2. Claudin-low and apocrine subtypes

Besides the five well-known intrinsic subtypes, microarray gene expression analyses have identified two more subtypes within ER-negative breast cancers: claudin-low and apocrine subtypes (Figure 1.2). The claudin-low subtype is non-basal, uncommon but an interesting subtype because it includes tumors that are characterized by the expression of genes associated with epithelial to mesenchymal-transition (EMT) such as low expression of *E-cadherin* and *claudins* and high expression *Vimentin* and stem cell markers (*$CD24^{low}CD44^{high}$* and *ALDH1*)[113]. The apocrine subtype includes ER-negative tumors that are also not of the basal-like subtype, but these tumors, however exhibit high expression of androgen receptor and other hormonally regulated genes (*SREBF1, CDH1, SCD, FASN, RND1,* and *GHR*)[114]. Patients with tumors displaying both the claudin-low and apocrine subtype features show poor overall survival compared to patients with the luminal subtype tumors[113].

1.6.3.3. Triple-negative breast cancer subtypes

Triple-negative breast cancer (TNBC) is a highly diverse group of breast cancers with varied response to the available chemotherapy. Recently, Lehmann and co-workers re-analyzed gene expression profiles of an extensive number of triple-negative breast tumors from 21 publically available data sets and identified 6 molecular subtypes within triple-negative disease[115]. These subtypes were characterized on the basis of gene ontology terms associated with differentially expressed genes and were named as: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM),

mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR). These subtypes exhibit distinct relapse free-survival (RFS), which was significantly lower in the LAR subtype compared with the BL1 subtype. The M subtype showed shorter RFS compared with BL1, while the MSL subtype had longer RFS than the M subtype. Microarray gene expression-based molecular subtypes including TNBC subtypes are schematically represented in Figure 1.3. It is worth mentioning however that the molecular classification of the breast cancer in general and TNBC in particular has led to the identification of active cellular pathways specific to each identified subtype and these pathways can contain potential target sites for personalized therapeutic strategies.

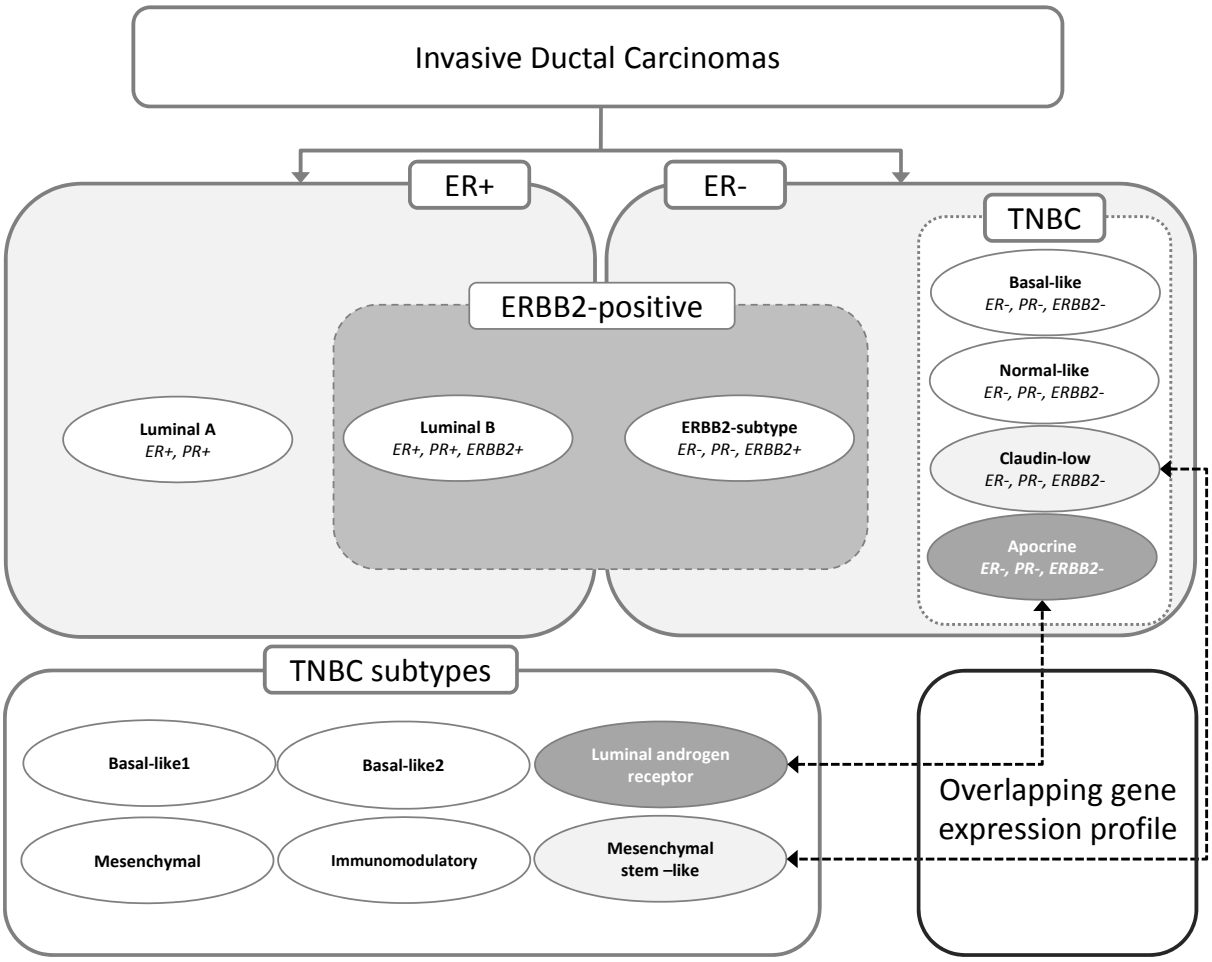


Figure 1.3 schematic representation of molecular subtypes of human breast cancer.

1.7 PROGNOSIS OF BREAST CANCER

Breast cancer is heterogeneous disease with respect to biological behavior and clinical outcome. Several prognostic markers have been identified and used in the clinic to accurately predict course of the disease in the patients. The currently available prognostic markers are generally divided into three classes:

1. Patient-related prognostic factors, such as age and menopausal status

2. Tumor-related prognostic factors, which include tumor size, tumor stage, axillary lymph node involvement, tumor differentiation status (grading), tumor ER, PGR, and ERBB2 expression status.
3. Molecular prognostic factors which include mRNA, miRNA, and protein expression signatures as well as epigenetic signatures including changes in DNA methylation and histone modification.

Clinical utility and significance of these prognostic markers have been extensively reviewed elsewhere[116-118].

1.8 BIOLOGY OF THE BREAST CANCER

Human mammary tissue is characterized by active cellular growth and proliferation. Mammary epithelial cells are organized in three dimensional structures, which are strongly dependent on a polarized morphology, specialized cell–cell contacts and specific attachments to an underlying basement membrane. These features are necessary for proper control of cell proliferation, survival, differentiation, migration and milk-protein secretion. Breast cancer develops through a multistep process in which cells become abnormal and multiply out of control. The whole process is governed by the genetic abrasions which occur in functionally important genes controlling the division of cells. Such mutated genes can be tumor suppressers and oncogenes. This escape can occur through the activation of a variety of molecular signaling pathways. In breast cancer, several biological signaling pathways are activated. These activated signaling pathways provide selective growth advantage to tumor cells over the normal cells located in the same tissue. Recently, non-coding RNAs such as microRNAs (miRNAs) also have emerged as important biological molecules that are involved in regulation of several genes involved in breast cancer initiation and progression[119-125].

1.8.1. MicroRNAs and breast cancer biology

MicroRNAs (miRNAs) are a class of 20-25 nucleotide-long non-coding RNAs that regulate gene expression through a complementary base-pairing with seed sequence present mostly in the 3' untranslated region (3’ UTR) of target mRNAs. miRNA biogenesis starts with RNA polymerase II mediated transcription of primary miRNA sequence (pri-miRNA). The pri-miRNA is then microprocessed to a 70-nucleotide stem loop known as the precursor miRNA (pre-miRNA) by the enzymatic activity of two endonucleases Drosha and DGCR8. The pre-miRNA is then actively transported to the cytoplasm via exportin 5 in a Ran-GTP-dependent manner. In the cytoplasm, the pre-miRNA is recognized by another RNase III endonuclease called Dicer, and is cleaved to generate a mature 20-nucleotide miRNA duplex[126]. Generally, only one strand is biologically selected and the other strand is degraded. The mature miRNA is loaded into RNA-induced silencing complex (RISC), which contains Argonaute (Ago) and a single stranded target RNA. The mature miRNA allows the RISC to recognize target mRNA via partial sequence complementarity present between the seed sequence of the miRNA and seed matching sequences of the target mRNA.

There are two known mechanisms by which the RISC can inhibit the expression of the target genes: 1- RISC removes the poly-A tail of the target mRNA (deadenylation) by fostering the activity of deadenylases and exposes mRNA for its nuclease mediated degradation, 2- RISC blocks the initiation of mRNA by either inhibiting the factors required to start mRNA initiation and elongation or by causing ribosome stalling. The RISC-bound mRNAs can be localized to sub-cytoplasmic compartments, known as P-bodies where they are either stored in reversible manner or degraded by the activity of Ago protein[127]

Over the past few years, miRNA profiling studies have led to the identification of several miRNAs that are aberrantly expressed in human breast cancer and much of research is focused on understanding their underlying biology. In breast and many other cancers, some miRNAs can function as tumor suppressors while other can act as oncogenes. Tumor formation may therefore occur by deletion of a tumor suppressor miRNAs and/or amplification or overexpression of an oncogenic miRNAs[128]. For instance, *hsa-miR-15*, *hsa-miR-16, hsa-let-7, hsa-miR-34b, hsa-miR-10b, hsa-miR-125b,* and *hsa-miR-145* are downregulated or deleted in leukemia, lung cancer and breast cancer[129-131] on the other hand *hsa-miR-155,* and *hsa-miR-21* are overexpressed or amplified in breast cancer[131,132]. Growing evidence demonstrated that miRNAs being novel modulators in cancer can also be efficiently therapeutically targeted, and consequently cancer progression can be reversed. For example, *hsa-miR-155* alone is sufficient to induce tumorigenesis, a systematic knockdown of *hsa-miR-155* as anti-tumor therapy may reduce tumorigenesis[133,134], similarly *hsa-miR-21* has been shown to increase metastatic ability of breast cancer cells and knockdown of which may lead to treat advanced breast cancer[135].

1.9 BREAST CANCER RESEARCH MODELS

To become a fully malignant cell, a mammary cell undergoes considerable amount of transformations at the genomic level. These changes enable the malignant cells to disentangle from their primary site, pass through the tumor basement membrane, enter into the circulation and disseminate to secondary sites to give rise to metastatic tumors. As a part of the metastatic process, the tumor cells are able to resist anoikis, evade immune recognition, and physical stress like low oxygen pressure and finally manage to coexist in a foreign microenvironment with the cells of different organs[136]. Every step of the metastatic process is triggered through inter- or/and intra-cellular communication involving expression modulation of many genes and proteins[136]. In fact, hundreds to a few thousand of genes have now been reported to be deregulated by genomic alteration in breast cancer[137,138]. The functional assessment of these genes is therefore imperative to gain biological insights, which will eventually help in designing targeted therapies for improved treatment. Successful unfolding of these processes in cancer biology will require appropriate study models that can mimic the real scenario of human malignant cells *in vivo*. Depending on the aspects to study in breast cancer, several model systems have been employed and scientific knowledge has been extrapolated to humans and then accordingly possible treatment options are explored. Although, it is unlikely that the complexity of cancer in human patients is entirely

modeled but by understanding the strengths and weakness of the available models, it is possible to choose an appropriate model to study an individual scientific problem and question. This holds true while using breast cancer cell lines and mice as models to understand breast carcinogenesis and to design preventive therapeutics.

1.9.1. Well-established human breast cancer cell lines

Breast cancer cell lines are renewable resources, which have been widely used for laboratory research particularly as an *in vitro* model for breast cancer. A significant amount of our current knowledge on breast cancer disease has been accumulated using breast cancer cell lines cultured in both *in vitro* and *in vivo* conditions. The fundamental aspects of breast carcinogenesis studied using breast cancer cell line models are growth factors-induced responses in malignant cells, evaluation of the functional role of clinically important genes through loss and gain of function studies, global gene expression profiling of pure tumor cells, and drug sensitivity studies. The first ever described human breast cancer cell line is BT20 that was established in 1958 from a patient who was treated in a hospital of Baltimore, (USA). Since then, more than 100 human breast cancer cell lines have been established and propagated as a continuous in vitro culture (Table 1.6). Despite of the fact many cell lines are successfully established, a handful number of them are used in studies. For example, MCF7, T-47D, MDA-MB-231, BT20, BT549, MDA-MD-468, SK-BR-3, SK-BR-7, ZR-75-1, and CAMA1 have frequently been reported as models to investigate breast cancer biology.

Cell lines have been generated from both primary tumors (BT20, BT474, BT549, and HCC-cell lines series) and metastasis (MDA-MB-231, MDA-MB-436, MDA-MD-468, and DU4475). The majority of the cell lines from metastatic origin have been isolated from pleural effusions and only some of them have been obtained from metastatic organs. For instance, MCF-7 has been isolated from pleural effusion and MDA-MB-361 and ZR-75-1 have been established from brain metastasis and ascitic fluid, respectively. Many breast cancer cell lines are typically placed into a group or as a series since they were isolated from the same patient or they were generated by one research group from different patients and propagated in similar laboratory conditions. Examples are the 'Hamon Cancer Center' (HCC) series, which comprises of 18 cell lines, which were successfully derived from a series of 177 tumor tissues.

A number of cell line variants have been generated from the most widely used cell lines to learn resistance and sensitivity level of clinically important genes. For instance, various clones of an anti-estrogen responsive ZR-75-1 cell lines showing varied anti-estrogen sensitivity level have been generated after transfection it with different constructs of *BCAR1* gene, a known tamoxifen resistance conferring gene. The ZR-75-1/BCAR1 is a completely resistant clone towards anti-estrogen ICI182,780 added at 100nm concentration in the culture medium, whereas ZR-75-1/BCAR1Δ135 and ZR-75-1/BCAR1ΔSH3 clones show partial resistance in the medium containing ICI182,780[139]. The second example is MDA-MB-231 cell line variants, which were isolated based on their ability to preferentially disseminate to different metastatic sites (bones, lungs, and brain)[111,140,141] and have different latency[142,143].

A particularly subset of tumors develops in women due to germline mutations in *BRCA1*, *BRCA2*, and *CHEK2* genes. It is however important to note that only limited or no corresponding cell lines are available to mimic biological characteristics of these germline tumors. For example, only HCC1937 cell lines have been derived from a primary tumor of a patient with known history of germline *BRCA1* mutation[144]. Mutational screening of 41 cell lines by our group, however, has revealed three additional breast cancer cell lines (MDA-MB-436, SUM149PT, and SUM1315MO2) with deleterious mutations in *BRCA1* gene and one cell line (UACC812) with the *CHEK2-1100delC* mutation[145]. This group of 4 *BRCA1* mutant cell lines can serve as a model to unravel underlying biology of germline *BRCA1* mutant breast cancers.

1.9.1.1. Breast cancer cell lines in 2-D and 3-D cultures

Breast cancer cell lines have conventionally been cultivated in two-dimensional (2-D) culture conditions to investigate global gene expression, global genetic profiling as well as breast cancer biology underlying cellular interactions involved in tumor-initiation, tumor cell proliferation, induction of apoptosis and migratory behavior during the progression of breast cancer, to study efficacy of the therapeutic agents, and many other applications[146-148]. In 2-D culture conditions, cells are grown on flat culture dishes made of either polystyrene plastic (Figure 1.4a). The principle limitation of 2-D culture systems, however, is that the cell lines propagated in these conditions markedly differ from the breast microenvironment and therefore the data obtained from these studies cannot always be linearly extrapolated to the clinical situation. Kenny and colleagues have developed a 3-D culture system in which cells are grown in 3-D reconstituted basement membrane cultures (Figure 1.4b)[149]. They compared molecular profiles of 25 breast cancer cell lines cultured in 2-D versus 3-D culture conditions. Though not surprisingly, molecular profiles of breast cancer cell lines were more similar to themselves than to other cells that were grown under the same culture conditions, suggesting that the 3-D culture environment does not promote global changes at the gene expression level [149]. However, a group of signal transduction genes were determined, which were significantly correlated with cells grown in 3-D culture conditions. The expression changes of these genes may accounts for the morphological and behavioral differences of the breast cancer cell lines grown in 3-D culture compared with 2-D cultures[149,150]. Importantly, critical signaling interactions between epithelial cells and stromal cells during breast cancer pathogenesis have been widely studied by culturing cells in 3-D culture conditions[51,152]. For example, the effect of stromal rigidity on tumor development and morphogenesis has elegantly been modeled by reducing the matrix compliance in 3-D culture system, which promotes tumor-like morphogenesis in normal mammary epithelial cells (Figure 1.4c). Altogether, these studies indicate that depending on requirement breast cancer cell lines can be used in both 2-D and 3-D cultures to investigate the complex nature of breast cancer development and progression.

1.9.1.2. Breast cancer cell lines in xenografts

Human breast cancer cell lines are grown as xenografts in animals such as mice and rats. The xenografts have generally been raised by injecting cancer cells into animals. One method involves

tail-vein injection of the tumor cells with the help of a fine needle (Figure 1.4d). This procedure is considered to be physiologically less relevant as it allows the tumor cells to colonize the organs directly without going through all metastatic steps commonly seen in patients such as invasion, extravasation, circulation, and colonization of organs. This method is however particularly suitable to investigate the preference of tumor cells to disseminate to various distant organ. A second method is subcutaneous injection of the tumor cells into the flank of the mouse (Figure 1.4d). The limitation of this method is that tumor cells are injected in a different type of microenvironment, which may alter the growth and metastatic potential of the implanted cells. The third and commonly preferred method is an orthotopic implantation of the tumor cells under the mammary fat pad of the mouse (Figure 1.4d). This method is thought to be physiologically most relevant to the human scenario, though, there are crucial differences between the mouse and human mammary stroma but it provides a more comparable microenvironment to that of human than subcutaneous implantation. Moreover, the implanted cells experience all essential steps necessary to establish metastasis in the distant organs including the complex tumor–stromal cell interactions that facilitate tumor formation and progression.

With regard to breast cancer cell line xenograft models, a fundamental question arises whether or not these xenograft experiments share any relevance to human breast cancer. This question has been elegantly addressed by Massague and colleague in 2005. They labeled MDA-MD-231 breast cancer cells with a reporter marker and injected them into the mouse body through tail-vein injection. Subsequently, they were able to select MDA-MD-231 variants that were preferentially metastasizing to lungs, bone, and brain[111,140,141]. By doing molecular profiling of these MDA-MD-231 variants, unique metastatic gene signatures were identified for bone, lungs, and brains relapse. Importantly, such signatures were recognized to have predictive value in patients who relapsed to bone, lungs, and brain[141]. All these studies together demonstrate relevance and utility of the xenograft assay in general and xenograft of breast cancer cell lines in particular to understand cancer initiation and metastasis.

1.9.1.3. Breast cancer cell lines as preclinical models

The anticancer activity of a drug has been evaluated by its ability to inhibit proliferation of a particular cancer cell typically growing in vitro conditions and this anticancer activity have often been predictive for breast cancer in the clinic. Human breast cancer cell lines have frequently been used as preclinical models both *in vitro* and *in vivo* as xenografts. Two recent studies, one by Garnett and coworker and another by Barretina and coworkers, have remarkably underscored the efficacy of cell lines as preclinical models[148,153]. These studies have used 40 and 56 breast cancer cell lines, respectively, to study the anti-cancer efficacy of many drugs. Additionally, several features including expression level, mutation status, deletion, and amplification of clinically important genes have been identified as robust predictors of sensitivity to the tested anticancer agents. For instance, activating mutations in *BRAF* and *NRAS* are predictors of sensitivity in models generated for MEK inhibitor PD-0325901. Other examples are *EGFR* mutation and *ERBB2* amplification/ overexpression, for erlotinib and lapatinib sensitivity, respectively [154]; *HGF* expression and *MET*

amplification for the MET/ALK inhibitor PF2341066[155]; and *MDM2* overexpression for Nutlin-2 sensitivity[148,156]. Though, cell lines grown in 2-D culture condition have provided initial landscapes of drug sensitivity patterns but it should be noted that such landscapes only reflect pharmacology of tumor cells and not tumor associated microenvironment and therefore improved models mimicking also the tumor microenvironment are desired.

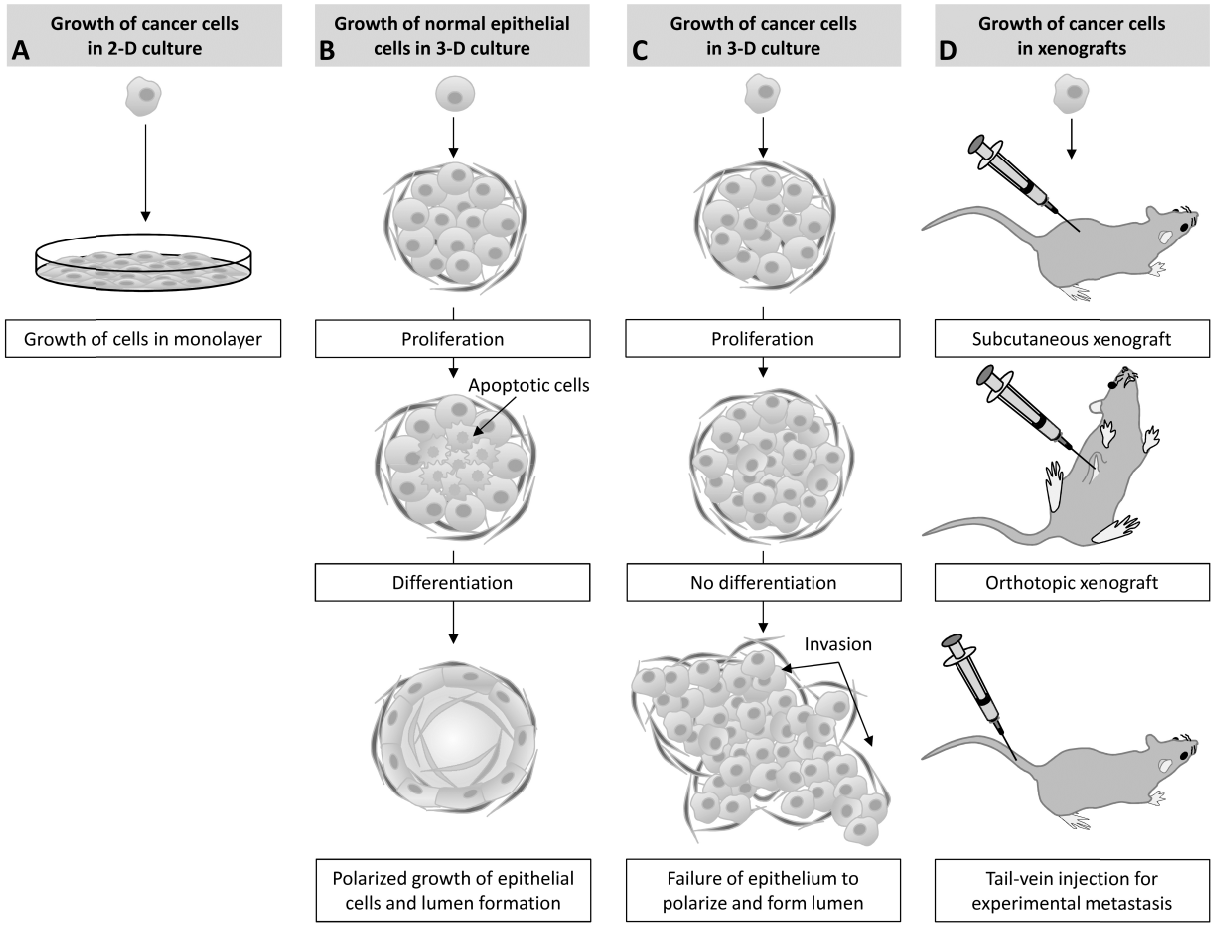


Figure 1.4 Cell culture and xenograft methods. Fundamental differences in the cell culture environments. **A**: Two-dimensional culture (2-D) grown on plastic lack exposure to the components of the extracellular matrix that are present in vivo. **B:** Morphology and behavior of non-transformed breast epithelial cells in three-dimensional (3-D) culture, which mimic more accurately breast structure and functions. Non-transformed cells grown in 3-D culture conditions undergo well-characterized morphogenesis processes. **C**: However, when breast cancer cells are grown in 3-D culture conditions they fail to become growth arrested compared to normal cells and they lack polarity and display aberrant architectures. **D**: Xenograft tumors are produced by injecting tumor cells subcutaneously into the flanks or orthotopically into the mammary fat pad of immunocompromised mouse. Xenograft tumors generally occur rapidly and reproducibly and large cohorts of tumor-bearing mice can be generated and used to examine anti-cancer activity of potential breast cancer drugs (pre-clinical drug screening). Xenografts have also been used to investigate metastasis processes. In spontaneous metastasis assays, metastasis arises from orthotopic tumors and tumor cells must compete in all steps of metastasis. In experimental metastasis assays, tumor cells are injected into the vasculature via intracardial or intravenous injection by bypassing some crucial steps of metastasis.

Table 1.6 Tumor origin, tumor type, and clinicopathological characteristics of well-established human breast cancer cell lines.

Cell lines	Origin	Tumor type	Chrom. number(s)	Sub-type	ER	PGR	ERBB2	TP53	PIK3CA	E-cad.	Ref.
SUM185PE	M (PE)	DC	NA	N	-	-	-	Mt	Mt	Wt	[157]
MDA-MB-175VII	M (PE)	IDC	49, 48	L	+	-	-	Wt	Wt	Wt	[158]
BT483	P	IDC	72	L	+	-	-	Mt	Mt	Wt	[159]
T47D	M (PE)	IDC	66	L	+	+	-	Mt	Mt	Wt	[160]
MDA-MB-415	M (PE)	AC	72	L	+	+	-	Mt	Wt	Wt	[161]
ZR75-1	M (As)	IDC	71-72	L	+	+	-	Wt	Wt	Wt	[162]
MCF-7	M (PE)	IDC	88	L	+	+	-	Wt	Mt	Wt	[163]
SUM52PE	NA	IDC	NA	L	+	+	-	Mt	Wt	Wt	[164]
UACC812	P	IDC	58-64	L	+	+	+	Wt	Wt	Wt	[165]
MDA-MB-361	M (Br)	AC	52	L	+	+	+	Mt	Mt	Wt	[161]
SUM190PT	P	IBC	NA	L	NA	NA	NA	Mt	Mt	Wt	[166]
BT474	P	IDC	55	L	-	-	+	Mt	Mt	Wt	[159]
MDA-MB-330	M (PE)	ILC	64	L	+	-	+	Mt	Wt	Wt	[161]
SUM225CWN	M	IDC	NA	L	-	-	+	Mt	Nd	Wt	[167]
UACC893	P	IDC	62	L	-	-	+	Mt	Mt	Wt	[165]
EVSA-T	M (As)	IDC	84	L	-	+	+	Mt	Wt	Mt	[168]
MDA-MB-453	M (PE)	AC	45	L	-	-	+	D	Mt	Mt	[161]
SK-BR-3	M (PE)	IDC	84	L	-	-	+	Mt	Wt	D	[166]
ZR75-30	M (As)	IDC	81	L	+	-	+	Wt	Wt	Mt	[162]
SK-BR-5	NA	NA	NA	L	-	-	+	Mt	Mt	Mt	[166]
OCUB-F, -M	M (PE)	RBC	84	L	-	-	+	Mt	Mt	D	[167]
MPE600	NA	IDC	46	L	+	-	+	Wt	Wt	D	[169]
MDA-MB-134VI	M (PE)	IDC	43	L	+	-	-	Mt	Wt	D	[158]
SUM44PE	M (PE)	HBC	60	L	+	+	-	Mt	Nd	Mt	[166]
CAMA-1	M (PE)	C	78	L	+	-	-	Mt	Wt	Mt	[170]
DU4475	M (Sk)	IDC	87-90	B	-	-	-	Wt	Wt	Wt	[171]
MDA-MB-468	M (PE)	AC	35	B	-	-	-	Mt	Wt	Wt	[172]
HCC1937	NA	NA	NA	L	-	-	-	Mt	Wt	Wt	[144]
BT20	P	IDC	49	C	-	-	-	Mt	Mt	Wt	[173]
SUM149PT	M (PE)	IBC	NA	C	-	-	-	Mt	Wt	H	[166]
SUM229PE	NA	NA	NA	L	-	-	-	Mt	Wt	H	[166]
SK-BR-7	NA	NA	NA	L	-	-	-	Wt	Wt	H	[166]
SUM102PT	P	IFL	48	Apo	NA	NA	NA	wt	Mt	H	[174]
Hs578T	P	CS	58	N	-	-	-	Mt	Wt	H	[175]
MDA-MB-231	M (PE)	IDC	64	N	-	-	-	Mt	Wt	H	[158]
SUM1315M02	NA	NA	NA	N	-	-	-	Mt	Wt	H	[166]
MDA-MB-436	M (PE)	AC	NA	N	-	-	-	Mt	Wt	H	[161]
BT549	P	PIDC	74	N	-	-	-	Mt	Wt	H	[159]
MDA-MB-157	M (PE)	MC	65	N	-	-	-	Mt	Wt	H	[176]
SUM159PT	NA	NA	NA	N	-	-	-	Mt	Mt	H	[164]
MDA-MB-435s	M (PE)	IDC	64	N	-	-	-	Mt	Wt	H	[161]
HCC1008	M (LN)	DC	NA	NA	NA	NA	NA	Mt	Wt	Wt	[177]
HCC1500	P	DC	NA	L	+	+	-	NA	Wt	Wt	[177]
HCC202	P	DC	NA	L	-	-	+	Mt	Wt	Wt	[177]
HCC1419	P	DC	NA	L	+	-	+	Mt	Wt	Wt	[177]
HCC2218	P	DC	NA	L	-	-	+	NA	Wt	Mt	[177]
HCC1569	P	MC	NA	N	-	-	+	Mt	Wt	Wt	[177]
HCC1954	M (PE)	AC	NA	C	-	-	+	Mt	Mt	Wt	[177]
HCC1599	P	DC	NA	C	-	-	-	Mt	Wt	Wt	[177]
HCC1806	P	DS	NA	C	-	-	-	Mt	Wt	Wt	[177]
HCC1187	P	DC	NA	C	-	-	-	Mt	Wt	Wt	[177]
HCC70	P	DC	NA	C	-	-	-	Mt	Wt	Wt	[177]

Table 1.6 Continued

Cell lines	Origin	Tumor type	Chrom. number(s)	Sub-type	ER	PGR	ERBB2	TP53	PIK3CA	E-cad.	Ref.
HCC1143	P	DC	NA	C	-	-	-	Mt	Wt	Wt	177
HCC1395	P	DC	NA	N	-	-	-	NA	Wt	Wt	177
HCC38	P	DC	NA	NA	NA	NA	NA	Mt	Wt	Wt	177
184A1	NA	RM	NA	B	NA	NA	NA	NA	NA	NA	178
CAL-51	M (PE)	AC	NA	B	NA	NA	NA	NA	NA	NA	179
EFM-19	M (PE)	IDC	NA	L	NA	NA	NA	NA	NA	NA	180
EFM-192A	M (PE)	AC	NA	L	NA	NA	+	NA	NA	NA	181
hTERT-HME1	NA	RM	NA	B	-	NA	-	NA	NA	NA	182
AU565	M (PE)	IDC	NA	NA	-	-	NA	NA	NA	NA	183
BOT-2	P	IDC	63	NA	NA	NA	NA	NA	NA	NA	184
BRC-230	P	IDC	60	NA	-	-	NA	NA	NA	NA	185
BrCa-MZ-01	P	MC	66	NA	+	+	NA	NA	NA	NA	186
BrCa-MZ-01	M (PE)	IBC	46	NA	-	-	NA	NA	NA	NA	186
BSMZ	M (PE)	IDC	80	NA	+	+	NA	NA	NA	NA	187
CAL-18A	P	C	71	NA	-	-	NA	NA	NA	NA	188
CAL-18B	P	C	65	NA	-	-	NA	NA	NA	NA	188
EP	M (PE)	IDC	53	NA	+	NA	NA	NA	NA	NA	189
GI-101	R (L)	IDC	98	NA	-	-	NA	NA	NA	NA	190
GCS	M (As)	IDC	NA	NA	+	+	NA	NA	NA	NA	191
HBL-100	NA	NA	63	B	-	-	NA	NA	NA	NA	192
HDQ-P1	P	IDC	55-59	NA	-	-	NA	NA	NA	NA	193
HH315	M (O)	C	113	NA	-	-	NA	NA	NA	NA	194
HH375	M (LN)	C	64	NA	-	-	NA	NA	NA	NA	194
Ia-270	M (PE)	IDC	NA	NA	+	+	NA	NA	NA	NA	195
IBEP-1	M (PE)	IDC	52	NA	-	+	NA	NA	NA	NA	196
IBEP-2	M (PE)	IDC	74	NA	+	-	NA	NA	NA	NA	196
IBEP-3	M (PE)	IDC	57	NA	-	+	NA	NA	NA	NA	196
IIB-BR-G	P	IDC	56	NA	-	-	NA	NA	NA	NA	197
JCK	M (PE)	IDC	NA	NA	+	+	NA	NA	NA	NA	191
KPL-1	M (PE)	IDC	77	NA	+	-	NA	NA	NA	NA	198
KPL-3C	M (PE)	IDC	63	NA	-	-	NA	NA	NA	NA	199
KPL-4	M (PE)	IDC	53	NA	-	-	NA	NA	NA	NA	200
LCC15-MB	M (F)	C	NA	NA	-	-	NA	NA	NA	NA	201
MA11	M (Bm)	ILC	64	NA	-	-	NA	NA	NA	NA	202
MAST	M (As)	IDC	60	NA	+	+	NA	NA	NA	NA	203
MaTu	M (LN)	IDC	66	NA	-	-	NA	NA	NA	NA	204
MFM-223	M (PE)	IDC	46	NA	-	-	NA	NA	NA	NA	205
MT-1	P (Tr)	C	46	NA	-	-	NA	NA	NA	NA	204
MT-3	P (Tr)	C	47	NA	-	-	NA	NA	NA	NA	204
MW	M (PE)	IDC	67	NA	-	NA	NA	NA	NA	NA	189
PMC42	M (PE)	C	66	NA	NA	NA	NA	NA	NA	NA	206
UISO-BCA-1	M (PE)	IDC	58	NA	-	-	NA	NA	NA	NA	207
UISO-BCA-2	M (PE)	IDC	62	NA	-	-	NA	NA	NA	NA	207
VHB-1	P	IDC	70	NA	+	+	NA	NA	NA	NA	208
MCF-10A	P	F	NA	B	-	-	NA	NA	NA	NA	209
MCF-12A	P	F	67	B	-	-	NA	NA	NA	NA	210

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; C: carcinoma; MC: medullary carcinoma; IBC: inflammatory breast carcinoma; PIDC: papillary invasive ductal carcinoma; AC: adenocarcinoma; CS: carcinoma sarcoma; F: fibrocystic disease; RM: reduction mammoplasty; P: primary tumor; M: metastasis; PE: pleural effusion; Br: brain; Sk: skin; Bm: bone marrow; CWN; chest wall nodule; As: ascites; Tr: transplanted; O: omenun; L: luminal type; B: basal-like; N: normal-like; Apo: apocrine subtype; Wt: wild-type; Mt: mutant; D: deleted; NA: not available; Chrom: chromosome; E-cad: E-cadherin; Ref: References

¶ER, PGR, and ERBB2 status expression is based on both protein and gene expression array data.

*Molecular subtype is based on microarray gene expression analysis.

1.9.2. Mice models for breast cancer study

Mice models have been extensively used to investigate fundamental biology underlying mammary development and tumorigenesis. The first category of mouse models includes mice in which mammary tumor development is induced chemically or by ionizing radiation. These mouse models are generally used to understand basic biological events underlying preliminary malignant lesions including ductal hyperplasia, atypical ductal hyperplasia, ductal carcinoma in-situ and their potential to develop invasive breast cancers[211,212]. The second category of the mouse models include genetically engineered mice (GEM) in which tumor development involves a comprehensive experimental driven genetic manipulation. The GEM models are typically used to unravel the biological roles of the tumor suppressor genes (*TP53*, *BRCA1, BRCA2*, *PTEN, CDH1*) and oncogenes (*PyMT*, *RAS,* c-*MYC*, *ERBB2*, *CCD1)*, which have been frequently implicated in breast cancer development and progression. GEM expressing oncogenes under the control of the MMTV-CK14 or WAP typically initiate tumors in the mammary glands, leading to primary tumors and then metastasis to distant organs during lateral stages of cancer progression[213]. Examples of GEM models are MMTV-PyMT, MMTV-Neu, MMTV-Neu^{NDL}, MMTV-Wnt1, and WAP-Ras[213-216].

Although, mice model studies have recapitulated the behavior and properties of human breast cancer tumors and revealed interaction between tumor cells and their microenvironment as well as explored molecular changes during metastasis, but data obtained from animal models as such is often not applicable to breast cancer patients due to several differences between human and mouse. For instance, no single mouse model fully recapitulates the molecular subtypes commonly present in human tumors and also very few ER-positive GEM exist. Furthermore, therapeutic strategies developed using GEM models could sometime lead to unexpected outcomes owing to fundamental differences in the components of the immune system[217].

1.10. SUMMARY

In summary, breast cancer is the most prevalent disease in the western world and a leading cause of cancer related deaths in females. The average lifetime risk of breast cancer varies between 9-12% all over the world. Major risk factors include female gender, age, endogenous and exogenous hormones, reproductive history, breast density, and a family history of cancer[17,18]. Breast cancer is a genetic disease of both the inherited and sporadic predisposition. Three well-defined groups of breast cancer susceptibility genes or loci have been identified: high-risk genes, moderate-risk genes, and low-risk loci, which altogether can explain about 35% of familial breast cancer risk. A vast majority of breast cancer is sporadic and somatic mutations conferring selective growth advantage on the cells are considered deriver mutations whereas somatic mutation with biologically neutral effect are considered passenger mutations. Historically, the breast cancer has been classified based on the presence of hormone receptors (ER and/or PGR). ER+ tumors accounts for about three quarters of the breast cancer and generally display better prognosis with targeted anti-hormonal therapeutic options. In contrast to it, ER- tumors are aggressive with worse outcome with no specific therapeutics options available. About 25-30% breast tumors have overexpression of ERBB2 protein and these tumors are treated with monoclonal antibodies such as trastuzumab.

Microarray-based gene expression profiling of the breast tumors have revealed several intrinsic molecular subtypes each with clinically distinct prognostic and prediction values. Biological complexity of the breast cancer is unfolded using many experimental models, among them human breast cancer cell lines and mouse models are more frequently applied models. More than 100 breast cancer cell lines have been established and cultured both *in vitro* under 2-D and 3-D culture conditions and *in vivo* as xenografts. Besides many possibilities of studying biology and genetics, breast cancer cell lines are also being frequently used as a preclinical model to screen potential breast cancer drugs. Moreover, several mouse models are being used to explore fundamental aspects of breast cancer carcinogenesis, especially, the onset of tumors and their metastasis to distant organs.

REFERENCES

1. Siegel, R., Ward, E., Brawley, O. & Jemal, A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* **61**, 212-36 (2011).
2. Forouzanfar, M.H. et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* **378**, 1461-84 (2011).
3. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60**, 277-300 (2010).
4. Louwman, W.J. et al. On the rising trends of incidence and prognosis for breast cancer patients diagnosed 1975-2004: a long-term population-based study in southeastern Netherlands. *Cancer Causes Control* **19**, 97-106 (2008).
5. Adriance, M.C., Inman, J.L., Petersen, O.W. & Bissell, M.J. Myoepithelial cells: good fences make good neighbors. *Breast Cancer Res* **7**, 190-7 (2005).
6. Gimpl, G. & Fahrenholz, F. The oxytocin receptor system: structure, function, and regulation. *Physiol Rev* **81**, 629-83 (2001).
7. Sopel, M. The myoepithelial cell: its role in normal mammary glands and breast cancer. *Folia Morphol (Warsz)* **69**, 1-14 (2010).
8. Keller, P.J. et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res* **12**, R87 (2010).
9. Haviv, I. Appraisal of progenitor markers in the context of molecular classification of breast cancers. *Breast Cancer Res* **13**, 102 (2011).
10. Gudjonsson, T. et al. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev* **16**, 693-706 (2002).
11. Shackleton, M. et al. Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84-8 (2006).
12. Stingl, J., Eaves, C.J., Zandieh, I. & Emerman, J.T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**, 93-109 (2001).
13. Lim, E. et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907-13 (2009).
14. Prat, A. & Perou, C.M. Mammary development meets cancer genomics. *Nat Med* **15**, 842-4 (2009).
15. Polyak, K. Breast cancer: origins and evolution. *J Clin Invest* **117**, 3155-63 (2007).
16. Ma, H., Bernstein, L., Pike, M.C. & Ursin, G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res* **8**, R43 (2006).

17. Garcia-Closas, M. et al. Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer* **95**, 123-9 (2006).
18. Sasieni, P.D., Shelton, J., Ormiston-Smith, N., Thomson, C.S. & Silcocks, P.B. What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries. *Br J Cancer* **105**, 460-5 (2011).
19. Althuis, M.D. et al. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiol Biomarkers Prev* **13**, 1558-68 (2004).
20. Huang, W.Y. et al. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. *Am J Epidemiol* **151**, 703-14 (2000).
21. Edge, S.B. & Compton, C.C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* **17**, 1471-4 (2010).
22. Elston, C.W. & Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-10 (1991).
23. Rakha, E.A. et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* **12**, 207 (2010).
24. Stratton, M.R. & Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**, 17-22 (2008).
25. Antoniou, A.C., Gayther, S.A., Stratton, J.F., Ponder, B.A. & Easton, D.F. Risk models for familial ovarian and breast cancer. *Genet Epidemiol* **18**, 173-90 (2000).
26. Antoniou, A.C. et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* **86**, 76-83 (2002).
27. Ghoussaini, M. et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-8 (2012).
28. Bradbury, A.R. & Olopade, O.I. Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* **8**, 255-67 (2007).
29. Masciari, S. et al. Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet* **44**, 726-31 (2007).
30. Schumacher, V. et al. STK11 genotyping and cancer risk in Peutz-Jeghers syndrome. *J Med Genet* **42**, 428-35 (2005).
31. Offit, K. et al. Germline BRCA1 185delAG mutations in Jewish women with breast cancer. *Lancet* **347**, 1643-5 (1996).
32. Antoniou, A.C. et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* **98**, 1457-66 (2008).
33. Hamel, N. et al. On the origin and diffusion of BRCA1 c.5266dupC (5382insC) in European populations. *Eur J Hum Genet* **19**, 300-6 (2011).
34. Neuhausen, S. et al. Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. *Nat Genet* **13**, 126-8 (1996).
35. Birch, J.M. et al. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* **20**, 4621-8 (2001).
36. Nelen, M.R. et al. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet* **13**, 114-6 (1996).
37. Lynch, E.D. et al. Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* **61**, 1254-60 (1997).
38. Jenne, D.E. et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* **18**, 38-43 (1998).
39. Berx, G., Becker, K.F., Hofler, H. & van Roy, F. Mutations of the human E-cadherin (CDH1) gene. *Hum Mutat* **12**, 226-37 (1998).
40. Heikkinen, K. et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis* **27**, 1593-9 (2006).
41. Hollestelle, A., Wasielewski, M., Martens, J.W. & Schutte, M. Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* **20**, 268-76 (2010).

42. Meijers-Heijboer, H. et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
43. Renwick, A. et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-5 (2006).
44. Seal, S. et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-41 (2006).
45. Rahman, N. et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**, 165-7 (2007).
46. Roeb, W., Higgins, J. & King, M.C. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. *Hum Mol Genet* **21**, 2738-44 (2012).
47. Goldgar, D.E. et al. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Res* **13**, R73 (2011).
48. Thorstenson, Y.R. et al. Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res* **63**, 3325-33 (2003).
49. Rafnar, T. et al. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet* **43**, 1104-7 (2011).
50. Tommiska, J. et al. Evaluation of RAD50 in familial breast cancer predisposition. *Int J Cancer* **118**, 2911-6 (2006).
51. Varon, R. et al. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467-76 (1998).
52. Pearson, T.A. & Manolio, T.A. How to interpret a genome-wide association study. *JAMA* **299**, 1335-44 (2008).
53. Christensen, K. & Murray, J.C. What genome-wide association studies can do for medicine. *N Engl J Med* **356**, 1094-7 (2007).
54. Erkko, H. et al. A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**, 316-9 (2007).
55. Mavaddat, N., Antoniou, A.C., Easton, D.F. & Garcia-Closas, M. Genetic susceptibility to breast cancer. *Mol Oncol* **4**, 174-91 (2010).
56. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
57. Ahmed, S. et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-90 (2009).
58. Cox, A. et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**, 352-8 (2007).
59. Stacey, S.N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-9 (2007).
60. Thomas, G. et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-84 (2009).
61. Stacey, S.N. et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-6 (2008).
62. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-8 (2009).
63. Turnbull, C. et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-7 (2010).
64. Antoniou, A.C. et al. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* **21**, 1-18 (2001).
65. Pharoah, P.D., Antoniou, A.C., Easton, D.F. & Ponder, B.A. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* **358**, 2796-803 (2008).
66. Devilee, P. & Rookus, M.A. A tiny step closer to personalized risk prediction for breast cancer. *N Engl J Med* **362**, 1043-5 (2010).
67. Garcia-Closas, M. et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**, e1000054 (2008).

68. Garcia-Closas, M. & Chanock, S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* **14**, 8000-9 (2008).
69. Hishikawa, Y., Tamaru, N., Ejima, K., Hayashi, T. & Koji, T. Expression of keratinocyte growth factor and its receptor in human breast cancer: its inhibitory role in the induction of apoptosis possibly through the overexpression of Bcl-2. *Arch Histol Cytol* **67**, 455-64 (2004).
70. Tamaru, N. et al. Estrogen receptor-associated expression of keratinocyte growth factor and its possible role in the inhibition of apoptosis in human breast cancer. *Lab Invest* **84**, 1460-71 (2004).
71. Zhang, Y. et al. Keratinocyte growth factor (KGF) induces aromatase activity in cultured MCF-7 human breast cancer cells. *Anticancer Res* **18**, 2541-6 (1998).
72. Zhang, Y. et al. Estrogen-induced keratinocyte growth factor mRNA expression in normal and cancerous human breast cells. *Oncol Rep* **5**, 577-83 (1998).
73. Zang, X.P. & Pento, J.T. Keratinocyte growth factor-induced motility of breast cancer cells. *Clin Exp Metastasis* **18**, 573-80 (2000).
74. Luqmani, Y.A., Graham, M. & Coombes, R.C. Expression of basic fibroblast growth factor, FGFR1 and FGFR2 in normal and malignant human breast, and comparison with other normal tissues. *Br J Cancer* **66**, 273-80 (1992).
75. Meyer, K.B. et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* **6**, e108 (2008).
76. Ahmadiyeh, N. et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A* **107**, 9742-6 (2010).
77. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).
78. Kan, Z. et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869-73 (2010).
79. Stratton, M.R. Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553-8 (2011).
80. Sjoblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-74 (2006).
81. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* (2012).
82. Wood, L.D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-13 (2007).
83. Shah, S.P. et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* (2012).
84. Copeland, N.G. & Jenkins, N.A. Deciphering the genetic landscape of cancer--from genes to pathways. *Trends Genet* **25**, 455-62 (2009).
85. Jonat, W. et al. Effectiveness of switching from adjuvant tamoxifen to anastrozole in postmenopausal women with hormone-sensitive early-stage breast cancer: a meta-analysis. *Lancet Oncol* **7**, 991-6 (2006).
86. Nilsson, S. & Gustafsson, J.A. Estrogen receptors: therapies targeted to receptor subtypes. *Clin Pharmacol Ther* **89**, 44-55 (2011).
87. Smith, I.E. & Dowsett, M. Aromatase inhibitors in breast cancer. *N Engl J Med* **348**, 2431-42 (2003).
88. Carey, L.A. et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**, 2492-502 (2006).
89. Vogel, C. et al. Management of ErbB2-positive breast cancer: insights from preclinical and clinical studies with lapatinib. *Jpn J Clin Oncol* **40**, 999-1013 (2010).
90. Sahin, A.A. Biologic and clinical significance of HER-2/neu (cerbB-2) in breast cancer. *Adv Anat Pathol* **7**, 158-66 (2000).

91. Sauter, G., Lee, J., Bartlett, J.M., Slamon, D.J. & Press, M.F. Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. *J Clin Oncol* **27**, 1323-33 (2009).
92. Slamon, D.J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-82 (1987).
93. Jones, C. et al. Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes. *Clin Cancer Res* **10**, 5988-97 (2004).
94. Sircoulomb, F. et al. Genome profiling of ERBB2-amplified breast cancers. *BMC Cancer* **10**, 539 (2010).
95. Geyer, C.E. et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* **355**, 2733-43 (2006).
96. Voduc, K.D. et al. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* **28**, 1684-91 (2010).
97. Hsieh, A.C. & Moasser, M.M. Targeting HER proteins in cancer therapy and the role of the non-target HER3. *Br J Cancer* **97**, 453-7 (2007).
98. Colombo, P.E., Milanezi, F., Weigelt, B. & Reis-Filho, J.S. Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. *Breast Cancer Res* **13**, 212 (2011).
99. Cheung, V.G. et al. Making and reading microarrays. *Nat Genet* **21**, 15-9 (1999).
100. Cooper, C.S. Applications of microarray technology in breast cancer research. *Breast Cancer Res* **3**, 158-75 (2001).
101. Reis-Filho, J.S. et al. Metaplastic breast carcinomas are basal-like tumours. *Histopathology* **49**, 10-21 (2006).
102. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747-52 (2000).
103. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-74 (2001).
104. Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**, 8418-23 (2003).
105. Yu, K., Lee, C.H., Tan, P.H. & Tan, P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. *Clin Cancer Res* **10**, 5508-17 (2004).
106. Han, W., Nicolau, M., Noh, D.Y. & Jeffrey, S.S. Characterization of molecular subtypes of Korean breast cancer: an ethnically and clinically distinct population. *Int J Oncol* **37**, 51-9 (2010).
107. Su, Y. et al. Distinct distribution and prognostic significance of molecular subtypes of breast cancer in Chinese women: a population-based cohort study. *BMC Cancer* **11**, 292 (2011).
108. Brenton, J.D., Carey, L.A., Ahmed, A.A. & Caldas, C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* **23**, 7350-60 (2005).
109. Dawood, S. et al. Defining breast cancer prognosis based on molecular phenotypes: results from a large cohort study. *Breast Cancer Res Treat* **126**, 185-92 (2011).
110. Smid, M. et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**, 3108-14 (2008).
111. Bos, P.D. et al. Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-9 (2009).
112. Kennecke, H. et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol* **28**, 3271-7 (2010).
113. Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* **12**, R68 (2010).
114. Farmer, P. et al. Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **24**, 4660-71 (2005).
115. Lehmann, B.D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750-67 (2011).

116. Reis-Filho, J.S. & Pusztai, L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* **378**, 1812-23 (2011).
117. Weigel, M.T. & Dowsett, M. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocr Relat Cancer* **17**, R245-62 (2010).
118. Taneja, P. et al. Classical and Novel Prognostic Markers for Breast Cancer and their Clinical Significance. *Clin Med Insights Oncol* **4**, 15-34 (2010).
119. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-66 (2006).
120. Saini, H.K., Griffiths-Jones, S. & Enright, A.J. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A* **104**, 17719-24 (2007).
121. Croce, C.M. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* **10**, 704-14 (2009).
122. Inui, M., Martello, G. & Piccolo, S. MicroRNA control of signal transduction. *Nat Rev Mol Cell Biol* **11**, 252-63 (2010).
123. Keklikoglou, I. et al. MicroRNA-520/373 family functions as a tumor suppressor in estrogen receptor negative breast cancer by targeting NF-kappaB and TGF-beta signaling pathways. *Oncogene* (2011).
124. Tavazoie, S.F. et al. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147-52 (2008).
125. Zahnow, C.A. & Baylin, S.B. Epigenetic networks and miRNAs in stem cells and cancer. *Mol Cell* **39**, 661-3 (2010).
126. Sontheimer, E.J. Assembly and function of RNA silencing complexes. *Nat Rev Mol Cell Biol* **6**, 127-38 (2005).
127. Rivas, F.V. et al. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**, 340-9 (2005).
128. O'Day, E. & Lal, A. MicroRNAs and their target gene networks in breast cancer. *Breast Cancer Res* **12**, 201 (2010).
129. Cimmino, A. et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944-9 (2005).
130. Johnson, S.M. et al. RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635-47 (2005).
131. Iorio, M.V. et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65**, 7065-70 (2005).
132. Zhu, S. et al. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res* **18**, 350-9 (2008).
133. Babar, I.A. et al. Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma. *Proc Natl Acad Sci U S A* (2012).
134. Kong, W. et al. MicroRNA-155 regulates cell survival, growth, and chemosensitivity by targeting FOXO3a in breast cancer. *J Biol Chem* **285**, 17869-79 (2010).
135. Yan, L.X. et al. Knockdown of miR-21 in human breast cancer cell lines inhibits proliferation, in vitro migration and in vivo tumor growth. *Breast Cancer Res* **13**, R2 (2011).
136. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
137. Fridlyand, J. et al. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* **6**, 96 (2006).
138. Pollack, J.R. et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* **99**, 12963-8 (2002).
139. Brinkman, A. et al. The substrate domain of BCAR1 is essential for anti-estrogen-resistant proliferation of human breast cancer cells. *Breast Cancer Res Treat* **120**, 401-8 (2010).
140. Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537-49 (2003).

141. Minn, A.J. et al. Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518-24 (2005).
142. Lu, T.L., Chang, J.L., Liang, C.C., You, L.R. & Chen, C.M. Tumor spectrum, tumor latency and tumor incidence of the Pten-deficient mice. *PLoS One* **2**, e1237 (2007).
143. Han, J.S. & Crowe, D.L. Steroid receptor coactivator 1 deficiency increases MMTV-neu mediated tumor latency and differentiation specific gene expression, decreases metastasis, and inhibits response to PPAR ligands. *BMC Cancer* **10**, 629 (2010).
144. Tomlinson, G.E. et al. Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. *Cancer Res* **58**, 3237-42 (1998).
145. Elstrodt, F. et al. BRCA1 mutation analysis of 41 human breast cancer cell lines reveals three new deleterious mutants. *Cancer Res* **66**, 41-5 (2006).
146. Lee, G.Y., Kenny, P.A., Lee, E.H. & Bissell, M.J. Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat Methods* **4**, 359-65 (2007).
147. Han, J. et al. Molecular predictors of 3D morphogenesis by breast cancer cell lines in 3D culture. *PLoS Comput Biol* **6**, e1000684 (2010).
148. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-7 (2012).
149. Kenny, P.A. et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol* **1**, 84-96 (2007).
150. Xu, F. & Burg, K.J. Three-dimensional polymeric systems for cancer cell studies. *Cytotechnology* **54**, 135-43 (2007).
151. Wang, X. et al. A complex 3D human tissue culture system based on mammary stromal cells and silk scaffolds for modeling breast morphogenesis and function. *Biomaterials* **31**, 3920-9 (2010).
152. Paszek, M.J. et al. Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241-54 (2005).
153. Garnett, M.J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570-5 (2012).
154. Konecny, G.E. et al. Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res* **66**, 1630-9 (2006).
155. Zou, H.Y. et al. An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer Res* **67**, 4408-17 (2007).
156. Muller, C.R. et al. Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A. *Int J Cancer* **121**, 199-205 (2007).
157. Powell, C.T. et al. Persistent membrane translocation of protein kinase C alpha during 12-0-tetradecanoylphorbol-13-acetate-induced apoptosis of LNCaP human prostate cancer cells. *Cell Growth Differ* **7**, 419-28 (1996).
158. Cailleau, R., Young, R., Olive, M. & Reeves, W.J., Jr. Breast tumor cell lines from pleural effusions. *J Natl Cancer Inst* **53**, 661-74 (1974).
159. Lasfargues, E.Y., Coutinho, W.G. & Redfield, E.S. Isolation of two human tumor epithelial cell lines from solid breast carcinomas. *J Natl Cancer Inst* **61**, 967-78 (1978).
160. Keydar, I. et al. Establishment and characterization of a cell line of human breast carcinoma origin. *Eur J Cancer* **15**, 659-70 (1979).
161. Brinkley, B.R. et al. Variations in cell form and cytoskeleton in human breast carcinoma cells in vitro. *Cancer Res* **40**, 3118-29 (1980).
162. Engel, L.W. et al. Establishment and characterization of three new continuous cell lines derived from human breast carcinomas. *Cancer Res* **38**, 3352-64 (1978).
163. Soule, H.D., Vazguez, J., Long, A., Albert, S. & Brennan, M. A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **51**, 1409-16 (1973).

164. Forozan, F. et al. Molecular cytogenetic analysis of 11 new breast cancer cell lines. *Br J Cancer* **81**, 1328-34 (1999).
165. Meltzer, P. et al. Establishment of two new cell lines derived from human breast carcinomas with HER-2/neu amplification. *Br J Cancer* **63**, 727-35 (1991).
166. Ethier, S.P., Mahacek, M.L., Gullick, W.J., Frank, T.S. & Weber, B.L. Differential isolation of normal luminal mammary epithelial cells and breast cancer cells from primary and metastatic sites using selective media. *Cancer Res* **53**, 627-35 (1993).
167. Sawada, T. et al. [Establishment and characterization of a human breast cancer cell line, OCUB-1]. *Hum Cell* **7**, 138-44 (1994).
168. Engel, L.W. & Young, N.A. Human breast carcinoma cells in continuous culture: a review. *Cancer Res* **38**, 4327-39 (1978).
169. Smith, H.S. et al. Immortalization in culture: occurrence at a late stage in the progression of breast cancer. *J Natl Cancer Inst* **78**, 611-5 (1987).
170. Dobrynin, Y.V. Establishment and Characteristics of Cell Strains from Some Epithelial Tumors of Human Origin. *J Natl Cancer Inst* **31**, 1173-95 (1963).
171. Langlois, A.J. et al. Morphological and biochemical properties of a new human breast cancer cell line. *Cancer Res* **39**, 2604-13 (1979).
172. Pathak, S., Siciliano, M.J., Cailleau, R., Wiseman, C.L. & Hsu, T.C. A human breast adenocarcinoma with chromosome and isoenzyme markers similar to those of the HeLa line. *J Natl Cancer Inst* **62**, 263-71 (1979).
173. Lasfargues, E.Y. & Ozzello, L. Cultivation of human breast carcinomas. *J Natl Cancer Inst* **21**, 1131-47 (1958).
174. Sartor, C.I., Dziubinski, M.L., Yu, C.L., Jove, R. & Ethier, S.P. Role of epidermal growth factor receptor and STAT-3 activation in autonomous proliferation of SUM-102PT human breast cancer cells. *Cancer Res* **57**, 978-87 (1997).
175. Hackett, A.J. et al. Two syngeneic cell lines from human breast tissue: the aneuploid mammary epithelial (Hs578T) and the diploid myoepithelial (Hs578Bst) cell lines. *J Natl Cancer Inst* **58**, 1795-806 (1977).
176. Young, R.K., Cailleau, R.M., Mackay, B. & Reeves, W.J., Jr. Establishment of epithelial cell line MDA-MB-157 from metastatic pleural effusion of human breast carcinoma. *In Vitro* **9**, 239-45 (1974).
177. Gazdar, A.F. et al. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. *Int J Cancer* **78**, 766-74 (1998).
178. Stampfer, M.R. & Bartley, J.C. Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A* **82**, 2394-8 (1985).
179. Gioanni, J. et al. Establishment and characterisation of a new tumorigenic cell line with a normal karyotype derived from a human breast adenocarcinoma. *Br J Cancer* **62**, 8-13 (1990).
180. Simon, W.E. et al. Alteration of steroid hormone sensitivity during the cultivation of human mammary carcinoma cells. *In Vitro* **20**, 157-66 (1984).
181. Uphoff, C.C., Denkmann, S.A., Steube, K.G. & Drexler, H.G. Detection of EBV, HBV, HCV, HIV-1, HTLV-I and -II, and SMRV in human and other primate cell lines. *J Biomed Biotechnol* **2010**, 904767 (2010).
182. Morales, C.P. et al. Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nat Genet* **21**, 115-8 (1999).
183. Bacus, S.S., Kiguchi, K., Chin, D., King, C.R. & Huberman, E. Differentiation of cultured human breast cancer cells (AU-565 and MCF-7) associated with loss of cell surface HER-2/neu antigen. *Mol Carcinog* **3**, 350-62 (1990).
184. Nordquist, R.E., Ishmael, D.R., Lovig, C.A., Hyder, D.M. & Hoge, A.F. The tissue culture and morphology of human breast tumor cell line BOT-2. *Cancer Res* **35**, 3100-5 (1975).

185. Amadori, D. et al. Establishment and characterization of a new cell line from primary human breast carcinoma. *Breast Cancer Res Treat* **28**, 251-60 (1993).
186. Mobus, V.J. et al. Differential characteristics of two new tumorigenic cell lines of human breast carcinoma origin. *Int J Cancer* **77**, 415-23 (1998).
187. Watanabe, M. et al. Establishment of the human BSMZ breast cancer cell line, which overexpresses the erbB-2 and c-myc genes. *Cancer Res* **52**, 5178-82 (1992).
188. Gioanni, J. et al. Establishment, characterization, chemosensitivity, and radiosensitivity of two different cell lines derived from a human breast cancer biopsy. *Cancer Res* **45**, 1246-58 (1985).
189. Chu, M.Y. et al. Differential characteristics of two newly established human breast carcinoma cell lines. *Cancer Res* **45**, 1357-66 (1985).
190. Hurst, J. et al. A novel model of a metastatic human breast tumour xenograft line. *Br J Cancer* **68**, 274-6 (1993).
191. Calaf, G. et al. New cell lines of human breast cancer origin. *Breast Cancer Res Treat* **21**, 63-75 (1992).
192. Gaffney, E.V. A cell line (HBL-100) established from human breast milk. *Cell Tissue Res* **227**, 563-8 (1982).
193. Wang, C.S. et al. Establishment and characterization of a new cell line derived from a human primary breast carcinoma. *Cancer Genet Cytogenet* **120**, 58-72 (2000).
194. Nayak, S.K. et al. Characterization of cancer cell lines established from two human metastatic breast cancers. *In Vitro Cell Dev Biol Anim* **36**, 188-93 (2000).
195. Loh, P.M. et al. Development of a new human breast cancer cell line Ia-270. *Breast Cancer Res Treat* **5**, 23-9 (1985).
196. Siwek, B., Larsimont, D., Lacroix, M. & Body, J.J. Establishment and characterization of three new breast-cancer cell lines. *Int J Cancer* **76**, 677-83 (1998).
197. Bover, L. et al. Description of a new human breast cancer cell line, IIB-BR-G, established from a primary undifferentiated tumor. *Breast Cancer Res Treat* **19**, 47-56 (1991).
198. Kurebayashi, J., Kurosumi, M. & Sonoo, H. A new human breast cancer cell line, KPL-1 secretes tumour-associated antigens and grows rapidly in female athymic nude mice. *Br J Cancer* **71**, 845-53 (1995).
199. Kurebayashi, J., Kurosumi, M. & Sonoo, H. A new human breast cancer cell line, KPL-3C, secretes parathyroid hormone-related protein and produces tumours associated with microcalcifications in nude mice. *Br J Cancer* **74**, 200-7 (1996).
200. Kurebayashi, J. et al. Isolation and characterization of a new human breast cancer cell line, KPL-4, expressing the Erb B family receptors and interleukin-6. *Br J Cancer* **79**, 707-17 (1999).
201. Thompson, E.W. et al. LCC15-MB: a vimentin-positive human breast cancer cell line from a femoral bone metastasis. *Clin Exp Metastasis* **17**, 193-204 (1999).
202. Rye, P.D. et al. Brain metastasis model in athymic nude mice using a novel MUC1-secreting human breast-cancer cell line, MA11. *Int J Cancer* **68**, 682-7 (1996).
203. Zoli, W. et al. A new cell line from human infiltrating ductal carcinoma of the breast: establishment and characterization. *J Cancer Res Clin Oncol* **122**, 237-42 (1996).
204. Hambly, R.J., Double, J.A., Thompson, M.J. & Bibby, M.C. Establishment and characterisation of new cell lines from human breast tumours initially established as tumour xenografts in NMRI nude mice. *Breast Cancer Res Treat* **43**, 247-58 (1997).
205. Hackenberg, R. et al. Androgen sensitivity of the new human breast cancer cell line MFM-223. *Cancer Res* **51**, 5722-7 (1991).
206. Whitehead, R.H., Monaghan, P., Webber, L.M., Bertoncello, I. & Vitali, A.A. A new human breast carcinoma cell line (PMC42) with stem cell characteristics. II. Characterization of cells growing as organoids. *J Natl Cancer Inst* **71**, 1193-203 (1983).
207. Mehta, R.R. et al. Human breast carcinoma cell lines: ultrastructural, genotypic, and immunocytochemical characterization. *Anticancer Res* **12**, 683-92 (1992).

208. Vandewalle, B. et al. Establishment and characterization of a new cell line (VHB-1) derived from a primary breast carcinoma. *J Cancer Res Clin Oncol* **113**, 550-8 (1987).
209. Soule, H.D. & McGrath, C.M. A simplified method for passage and long-term growth of human mammary epithelial cells. *In Vitro Cell Dev Biol* **22**, 6-12 (1986).
210. Paine, T.M., Soule, H.D., Pauley, R.J. & Dawson, P.J. Characterization of epithelial phenotypes in mortal and immortal human breast cells. *Int J Cancer* **50**, 463-73 (1992).
211. Medina, D. & Smith, G.H. Chemical carcinogen-induced tumorigenesis in parous, involuted mouse mammary glands. *J Natl Cancer Inst* **91**, 967-9 (1999).
212. Hennighausen, L. Mouse models for breast cancer. *Breast Cancer Res* **2**, 2-7 (2000).
213. Maglione, J.E. et al. Transgenic Polyoma middle-T mice model premalignant mammary disease. *Cancer Res* **61**, 8298-305 (2001).
214. Bouchard, L., Lamarre, L., Tremblay, P.J. & Jolicoeur, P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-neu oncogene. *Cell* **57**, 931-6 (1989).
215. Siegel, P.M., Dankort, D.L., Hardy, W.R. & Muller, W.J. Novel activating mutations in the neu proto-oncogene involved in induction of mammary tumors. *Mol Cell Biol* **14**, 7068-77 (1994).
216. Jonkers, J. & Derksen, P.W. Modeling metastatic breast cancer in mice. *J Mammary Gland Biol Neoplasia* **12**, 191-203 (2007).
217. Sharpless, N.E. & Depinho, R.A. The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat Rev Drug Discov* **5**, 741-54 (2006).

Chapter 2

Aims and outlines of the thesis

AIMS AND OUTLINES OF THE THESIS

Breast cancer, a world-wide leading cause of cancer deaths, is initiated by aberrations arising in the genome of tumor initiating cells that line the ducto-lobular epithelium of breast mammary tissue[1-3]. These aberrations disrupt normal biological signaling taking place amongst epithelial cells and between epithelial cells and surrounding stromal cells. This aberrant signaling leads to uncontrolled growth of these tumor initiating epithelial cells, which subsequently acquire additional genetic and epigenetic modifications driving aberrant expression of numerous crucial regulatory molecules including transcription factors and microRNAs and therefore might become highly aggressive and refractory to the currently available therapeutics. Genetic aberrations driving the initiation of breast cancer appear, depending on the risk they confer, in three flavors: highly penetrant genes, intermediate, and low-risk genes. Successful management of breast cancer can be improved, firstly, by performing systematic studies on the biological consequences of known genetic variants and molecules involved in biologically important signaling pathways, secondly, by discovering novel genetic variants causal to breast cancer susceptibility, and thirdly, by determining novel markers, which can accurately predict the course of disease in patients. Finally, to better study and understand the biological consequences of revealed genetic drivers the development of new model systems and thorough characterization of *in vitro* and *in vivo* model systems is also of utmost importance.

The study described in this thesis focused on two topics:

1. The discovery of novel susceptibility loci and the determination of the clinical and biological significance of the newly discovered low-risk breast cancer susceptibility loci.
2. The characterization of current *in vitro* experimental model system and the development of *in vivo* xenografts models.

After the discovery of CHEK2 1100delC as a novel polygenic intermediate risk allele for hereditary breast and colorectal cancer (HBCC) families[4-5], we aimed in **chapter 3** to find additional polygenic drivers for these families. As germ-line mutations in MSH6 predispose to Lynch syndrome and as previous work had shown, in a limited number of cases, the co-existence of CHEK2 1100delC and MSH6 mutations, we therefore focused on genetic variants in MSH6 gene. In **chapters 4-5**, the newly discovered low-risk breast cancer susceptibility alleles rs2981282 (*FGFR2*), rs3817198 *(LSP1)*, rs889312 *(MAP3K1)*, rs3803662 (*TOX3* or *TNRC9*), rs2107425 *(H19),* rs13281615 (*POUF3*/*MYC*), and rs13387042 (2q35) were investigated. These SNPs, despite of being very common in the general population only confer relative to *BRCA1* and even *CHEK2* a modest increase of risk to virtually all breast cancers[6-7]. As most SNPs do not alter the coding sequence and are often not even located in the vicinity of genes their mode of action is poorly understood. It was hypothesized that the low-risk SNPs modulate the expression levels of nearby located genes[8]. To shed light on this we determined minor allele frequency (MAF) of these low-risk breast cancer loci in both breast cancer

cell lines (**chapter 4**) and primary breast tumors (**chapter 5**) and measured the mRNA expression levels of the corresponding genes located nearby in the genome. In addition, as for the primary breast tumors investigated (**chapter 5**) we also had extensive clinical follow-up, we therefore determined if these SNPs contribute to disease progression beyond the stage of tumor initiation by evaluating their prognostic significance. To further understand breast cancer progression, in **chapter 6** we investigated in primary breast cancer patients the prognostic and biological significance of TWIST1, an emerged marker for epithelial to mesenchymal transition (EMT). For this, *TWIST1* mRNA expression was determined in a cohort of 1,427 primary breast cancer patients with extensive clinical follow-up to determine its association with disease progression and to reveal whether in clinical breast cancer specimen *TWIST1* also related to EMT. For the latter, we extracted TWIST1's biological context *in silico* from co-expressed genes measured on gene expression arrays. In **chapter 7 and 8**, we focused on our large collection of 51 well-characterized human breast cancer cell lines to find further support whether these cell lines represent clinical breast tumors (**Chapter 8**) and, to further explore their intrinsic and genetic diversity (**Chapter 7**). For this, we performed genome-wide profiling, at the mRNA and miRNA transcriptome levels of this entire panel of cell lines in which the genetic mutation diversity had previously been characterized by sequencing the most common tumor suppressor- and oncogenes[9]. In **chapter 7**, we associated the intrinsic subtypes and common genetic variation with the miRNA transcriptome. In **chapter 8**, we developed in mice preclinical xenograft models equipped with a luciferase reporter gene to allow *in vivo* imaging. Such models can be used to gain biological insights of breast cancer biology *in vivo* and they might also have potential use as preclinical xenograft model for *in vivo* drug screening. For this thesis we selected multiple cell lines from each intrinsic subtype to address whether these models recapitulate the *in vivo* observed diversity among subtypes with regard to growth characteristics and metastatic behavior. The results of the experiments performed in **chapter 3 to 8** and their clinical and biological implications are discussed in **chapter 9** and are summarized in **chapter 10**.

REFERENCES

1. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60**, 277-300 (2010).
2. Siegel, R., Ward, E., Brawley, O. & Jemal, A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* **61**, 212-36 (2011).
3. Stephens, P.J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-4 (2012).
4. Meijers-Heijboer, H. et al. The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet* **72**, 1308-14 (2003).
5. Meijers-Heijboer, H. et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
6. Hindorff, L.A., Gillanders, E.M. & Manolio, T.A. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis* **32**, 945-54 (2011).

7. Ghoussaini, M. et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-8 (2012).
8. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
9. Hollestelle, A. et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* **121**, 53-64 (2010).

Chapter 3

Association of rare MSH6 variants with familial breast cancer

Marijke Wasielewski, **Muhammad Riaz**, Joyce Vermeulen, Ans van den Ouweland, Ineke Labrijn-Marks, Renske Olmer, Linda van der Spaa, Jan G.M. Klijn, Hanne Meijers-Heijboer, Dennis Dooijes, and Mieke Schutte.
Breast Cancer Res Treat 2010; 123(2):315-20.

ABSTRACT

Germline mutations in the mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2* predispose to Lynch syndrome (also known as hereditary non-polyposis colorectal cancer). Recently, we have shown that the *CHEK2* 1100delC mutation also is associated with Lynch syndrome/Lynch syndrome-associated families albeit in a polygenic setting. Two of the ten *CHEK2* 1100delC positive Lynch syndrome families additionally carried a pathogenic *MLH1* or *MSH6* mutation, suggesting that mutations in mismatch repair genes may be involved in *CHEK2* 1100delC associated cancer phenotypes. A phenotype of importance is hereditary breast and colorectal cancer (HBCC), with the *CHEK2* 1100delC mutation present in almost one-fifth of the families – again in a polygenic setting. To evaluate the involvement of *MSH6* in polygenic CHEK2 cancer susceptibility, we here have analyzed the entire *MSH6* coding sequence for genetic alterations in 68 HBCC breast cancer families. Rare *MSH6* variants, with population frequencies below 1%, were identified in 11.8% of HBCC breast cancer families whereas the same variants were identified in only 1.5% of population controls, suggesting that rare *MSH6* variants are associated with HBCC breast cancer ($P \leq 0.00001$). However, screening of the entire *MSH6* coding sequence in 68 non-HBCC breast cancer families showed a similar association (8.8% vs. ~1. 4% in controls, $P \leq 0.001$), suggesting that rare *MSH6* variants are not confined to HBCC breast cancer. Together, our data suggest that rare *MSH6* variants may predispose to familial breast cancer. However, none of the rare *MSH6* variants are obviously pathogenic, suggesting that a more subtle disease mechanism may operate in breast carcinogenesis.

INTRODUCTION

In 2002, we and others have identified the truncating *CHEK2* 1100delC mutation as the first moderate-risk breast cancer susceptibility allele, present in about 5% of Dutch breast cancer families[1,2]. In addition, we have shown that the *CHEK2* 1100delC mutation was also present among Lynch syndrome/Lynch syndrome-associated families and was particularly prevalent among breast cancer families with a hereditary breast and colorectal cancer (HBCC) phenotype (4% and 18%, respectively)[3,4]. However, in each of these instances, *CHEK2* 1100delC appeared to confer cancer risks in a polygenic setting.

The association of *CHEK2* 1100delC with colorectal cancer phenotypes suggested known colorectal cancer genes as likely candidates for polygenic CHEK2 cancer susceptibility[4]. Germline mutations in the DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* and *PMS2* have been identified as the major causes for Lynch syndrome (previously also known as hereditary non-polyposis colorectal cancer, HNPCC[5-7]. Interestingly, in two of the ten *CHEK2* 1100delC-positive Lynch syndrome-associated families we had also identified pathogenic mutations in *MHL1* or *MSH6*[4]. Concomitance of *CHEK2* and *MSH6* mutations had reportedly also been identified in a Finnish breast cancer family and in the colorectal cancer cell line HCT15[8-11]. During repair of DNA damage, activated CHEK2 is known to signal to BRCA1, which in turn acts as a scaffold protein for several DNA damage response proteins, including MSH6[12-14]. The close functional relation of CHEK2 and MSH6, together with the observed concomitance of mutations in the two genes suggested that *MSH6* mutations may be involved in CHEK2 polygenic cancer susceptibility. This hypothesis was particular appealing for the HBCC phenotype as *CHEK2* 1100delC is prevalent among these families whereas *MSH6* mutations are identified among Lynch syndrome families that also include breast cancer[3,15-17]. Here, we therefore have screened the *MSH6* gene for genetic alterations in 68 HBCC and 68 non-HBCC breast cancer families.

PATIENTS AND METHODS

Breast cancer families and population controls

One hundred thirty-six breast cancer families were selected from 578 breast cancer families registered at the Rotterdam Family Cancer Clinic at Erasmus MC. Breast cancer families were classified as hereditary breast and colorectal cancer (HBCC) families (n=68) or non-HBCC families (n=68). All families included at least two first or second-degree relatives (DGRs) with breast cancer of whom at least one was diagnosed before the age of 60 years. HBCC families additionally included at least one patient with breast and colorectal cancer *or* at least one patient with colorectal cancer diagnosed before the age of 50 years who is within second DGR of a breast cancer patient *or* at least two colorectal cancer patients of whom at least one is within second DGR of a breast cancer patient[3].

All breast cancer families were screened for germline mutations in the *BRCA1* and *BRCA2* genes and the *CHEK2* 1100delC mutation, identifying 26 *BRCA1* mutant families, 6 *BRCA2* mutant families and

18 families with the *CHEK2* 1100delC mutation. Two families were double mutant for *BRCA1* and the *CHEK2* 1100delC mutation. As part of our matching procedure, mutant *BRCA1*, *BRCA2* and *CHEK2* 1100delC families were equally divided over HBCC and non-HBCC families. Pathogenic germline mutations in the *MLH1* and *MSH2* gene were not identified in any of the 68 HBCC families. *MLH1* and *MSH2* mutation status had not been determined for the 68 non-HBCC families.
All breast cancer families originated from the southwestern Netherlands and have consented to search for cancer susceptibility genes. The 166 control individuals were geographically-matched to the familial breast cancer cases and included spouses of heterozygous carriers of cystic fibrosis gene mutations, ascertained through the department of Clinical Genetics at Erasmus MC. The medical ethical committee of Erasmus MC approved this study.

MSH6 mutation analysis

The mismatch repair gene *MSH6* (NM_000179.2) was screened for mutations in blood-derived DNA of the youngest diagnosed breast cancer case in the family (index case). The complete coding sequence of the *MSH6* gene, including intron/exon boundaries up to 25 bases in the intron, was amplified by standard PCR. Amplified fragments were subsequently analyzed for genetic alterations by *denaturing gradient gel electrophoresis* (*DGGE*) or by direct sequencing as described[17-19]. Unique sequence alterations were confirmed at least once by sequence analysis of an independently generated PCR product. Primer sequences and reaction conditions are available upon request.

Statistical analysis

The difference between the mutation frequency in breast cancer patients versus controls was analysed using Fisher's Exact Test. *P*-values of 0.05 or smaller were considered significant. All statistical analyses were performed with STATA statistical package, release 10 (STATA Corp, College Station, TX).

RESULTS AND DISCUSSION

Rare *MSH6* variants associate with HBCC breast cancer

We analyzed the entire *MSH6* coding sequence in 68 breast cancer families with hereditary breast and colorectal cancer (HBCC). Sequence analysis identified 25 different sequence alterations among 68 HBCC families, including 10 intronic and 15 exonic variants (Table 3.1). To evaluate the significance of the identified *MSH6* variants for HBCC, all variants were subsequently genotyped in 166 geographically-matched controls. Rare variants, with population frequencies below 1%, represented 56% (14/25) of the *MSH6* variants and each of these were identified only once or twice among the HBCC cohort. Four of the fourteen rare *MSH6* variants identified among HBCC families were also identified in the control cohort. Thus, although none of the rare *MSH6* variants associated significantly with HBCC families, the combined frequency of the rare variants was significantly higher among HBCC families compared to controls (16/136 (11.8%) vs. 5/306-332 (~1.5%), *P*

<0.00001; Table 3.1). These results suggested that rare *MSH6* variants may predispose for HBCC breast cancer.

Table 3.1 *MSH6* mutation analysis among 68 HBCC and 68 non-HBCC breast cancer families

MSH6 gene sequence[a]	Predicted protein effect	Minor allele/Total tested alleles (%)			
		HBCC	Non-HBCC	All BRC	Controls
Prevalent variants (>1%)					
c.260 +22C > G	–	31/136 (22.8)	23/136 (16.9)	54/272 (19.9)	58/332 (17.5)
c.3438 + 14A > T	–	58/136 (42.6)	58/136 (42.6)	116/272 (42.6)	126/322 (39.1)
c.3557-4dupT	–	26/136 (19.1)	26/136 (19.1)	52/272 (19.1)	56/324 (17.3)
c.4002-10delT	–	35/136 (25.7)	39/136 (28.7)	74/272 (27.2)	2/328 (25.6)
c.4002-10dupT	–	14/136 (10.3)	16/136 (11.8)	30/272 (11.0)	41/312 (13.1)
c.116G > A	p.G39E	25/136 (18.4)	24/136 (17.6)	49/272 (18.0)	52/332 (15.7)
c.186C > A	p.=	31/136 (22.8)	23/136 (16.9)	54/272 (19.9)	58/332 (17.5)
c.276A > G	p.=	29/136 (21.3)	22/136 (16.2)	51/272 (18.9)	53/332 (16.0)
c.540T > C	p.=	42/136 (30.9)	39/136 (28.7)	81/272 (29.8)	95/332 (28.6)
c.642C > T	p.=	12/136 (8.8)	14/136 (10.3)	26/272 (9.6)	38/332 (11.4)
c.1286C > G	p.L396V	2/136 (1.5)	0/136 –	2/272 (0.7)	6/332 (1.8)
Rare variants (< 1%)					
Intronic (up to 25 bases)					
c.457 + 13A > G	–	1/136 (0.7)	0/136 –	1/272 (0.4)	1/332 (0.3)
c.3439-16C > T	–	2/136 (1.5)	0/136 –	1/272 (0.4)	1/332 (0.3)
c.3647-11T > G	–	0/136 –	1/136 (0.7)	1/272 (0.4)	0/254 –
c.4001+12_4001+15delACTC	–	2/136 (1.5)	1/136 (0.7)	3/272 (1.1)	2/328 (0.6)
c.4002-10delTT	–	1/136 (0.7)	0/136 –	1/272 (0.4)	0/312 –
c.4002-10delTTT	–	0/136 –	1/136 (0.7)	1/272 (0.4)	1/312 (0.3)
c.4002-10delTTTT	–	0/136 –	1/136 (0.7)	1/272 (0.4)	0/312 –
c.4002-10dupTTT	–	1/136 (0.7)	1/136 (0.7)	2/272 (0.7)	0/312 –
Exonic synonymous					
c.1053C > T	p.=	1/136 (0.7)	0/136 –	1/272 (0.4)	0/332 –
c.1164C> T	p.=	0/136 –	1/136 (0.7)	1/272 (0.4)	0/332 –
c.2272C > T	p.=	0/136 –	1/136 (0.7)	1/272 (0.4)	0/332 –
c.2775A > C	p.=	1/136 (0.7)	0/136 –	1/272 (0.4)	0/332 –
c.3246G > T	p.=	1/136 (0.7)	2/136 (1.5)	3/272 (1.1)	0/332 –
c.3306T > A	p.=	1/136 (0.7)	1/136 (0.7)	2/272 (0.7)	0/332 –
Exonic non-synonymous					
c.73G > T	p.A25S	1/136 (0.7)	1/136 (0.7)	2/272 (0.7)	1/332 (0.3)
c.751A > G	p.I251 V	1/136 (0.7)	0/136 –	1/272 (0.4)	0/332 –
c.1508C> G	p.S503C	1/136 (0.7)	0/136 –	1/272 (0.4)	1/332 (0.3)
c.2045C> T	p.S682F	1/136 (0.7)	0/136 –	1/272 (0.4)	0/332 –
c.2633T > C	p.V878A	1/136 (0.7)	0/136 –	1/272 (0.4)	0/332 –
c.4072_4075dupGATT	p.K1358fsX2[b]	0/136 –	1/136 (0.7)	1/272 (0.4)	0/332 –

Table 3.1 continued (Summary rare variants)

	Minor alleles/Total tested alleles (%)				
	All	**Intronic**	**Exonic all**	**Exonic synonymous**	**Exonic non-synonymous**
HBCC families	16/136 (11.8)	7/136 (5.1)	9/136 (6.6)	4/136 (2.9)	5/136 (3.7)
Controls	5/306–332 (~1.5)	3/306–332 (~0.9)	2/332 (0.6)	0/332 (~0.0)	2/332–332 (0.6)
P value	< 0.00001				
Non-HBCC families	12/136 (8.8)	5/136 (3.7)	7/136 (5.1)	5/136 (3.7)	2/136 (1.5)
Controls	4/254–332 (~1.4)	3/254–328 (~1.1)	0/332 (~0.0)	0/332 (~0.0)	1/332 (0.3)
P value	<0.001				
All BRC families	28/272 (10.3)	12/272 (4.4)	16/272 (5.9)	9/272 (3.3)	7/272 (2.6)
Controls	6/254–332 (~2.1)	4/254–332 (~1.4)	2/332 (0.6)	0/332 (~0.0)	2/332 (0.6)
P value	<0.0001				

[a] Numbering of nucleotide changes according to *MSH6* GenBank sequence NM_000179.2
[b] Frame shift mutation is indicated by the first changed codon and the number of newly encoded codons, including premature termination codon X
This particular variant locates 30 in the gene sequence predicting a protein shortened with only two amino acids. BRC breast cancer, HBCC hereditary breast and colorectal cancer

Rare *MSH6* variants associate with familial breast cancer

To evaluate if rare *MSH6* variants associate with HBCC breast cancer or rather with familial breast cancer in general, we also screened the entire *MSH6* coding sequence in 68 matched non-HBCC families (Table 3.1). Non-HBCC families were matched to the HBCC families with respect to their cancer pattern, i.e. the number of cancer patients per family, the number of cancer patients overall and the number of patients with single, double, triple or more cancers (Table 3.2). To avoid selection for HBCC, the non-HBCC families have been counter-selected for the presence of colorectal cancer. Six additional rare *MSH6* variants were identified among the non-HBCC families of which only one was also identified among the controls (Table 3.1). The combined frequency of rare *MSH6* variants was also significantly higher among non-HBCC families compared to controls (12/136 (8.8%) vs. 4/254-332 (~1.4%), *P* <0.001; Table 3.1). Hence, the prevalence of rare *MSH6* variants among non-HBCC families is equally high as among HBCC families, indicating that rare *MSH6* variants are not only associated with HBCC breast cancer but more likely with breast cancer in general. Combining the data on HBCC and non-HBCC families revealed a prevalence of rare *MSH6* variants of 10.3% among all breast cancer families whereas the same variants were identified in approximately 2.1% of controls (*P* <0.0001; Table 3.1), strongly suggesting that rare *MSH6* variants may predispose for familial breast cancer.

Do rare *MSH6* variants predispose for familial breast cancer?

Two aspects may question whether rare *MSH6* variants indeed predispose for familial breast cancer. First, we have not screened the entire *MSH6* coding sequence in the control cohort. Our strategy had been to genotype in the controls only those *MSH6* variants that had been identified among the familial breast cancer cohorts. Apart from the genotyped variant sequence, our primer design allowed analysis of the sequence surrounding the genotyped variant often including entire exon sequences and intron sequences up to 25 base pairs. In this way we were able to analyse 75% of

the *MSH6* coding sequence in all controls. In addition to the five rare *MSH6* variants that already had been identified in the breast cancer families, two variants (c.59C>T, p.A20V and c.4002-10delCT) were once identified exclusively among the controls. Based on this low prevalence of rare *MSH6* variants that are exclusively present among the controls (2/254 (0.8%); Table 3.1), it may be anticipated that screening of the remaining quarter of the *MSH6* coding sequence in the controls is unlikely to identify many more rare *MSH6* variants.

Another aspect is the underlying disease mechanism of how rare *MSH6* variants may predispose to breast cancer. We found that the prevalence of rare *MSH6* variants was consistently higher among the breast cancer families than the controls, whether they were intronic or exonic and whether they were synonymous or non-synonymous. However, none of the identified rare *MSH6* variants are known pathogenic mutations. Evaluation of each of the rare *MSH6* variants with the splice prediction programs Berkeley Drosophila Genome Project (www.fruitfly.org/seq_tools/splice.html) and NetGene2 (www.cbs.dtu.dk./services/NetGene2/) did not predict structural effects. It thus remains unclear how these rare *MSH6* variants exert their putative oncogenic effect, particularly for the intronic variants and the non-synonymous exonic variants. But then, are obvious pathogenic effects to be expected in a polygenic setting? Importantly, the statistical evidence from our analysis that rare *MSH6* variants associate with familial breast cancer is compelling ($P<0.001$), strongly suggesting cancer predisposition by rare *MSH6* variants. Obviously, our analysis requires replication in independent familial breast cancer cohorts, and it seems also warranted to screen familial colorectal cancer cohorts, with particular focus to rare variants that are not obviously pathogenic. In this respect, our current findings are consistent with a report by Nevanlinna and colleagues in which 15 different *MSH6* variants were identified among 38 breast cancer families with colorectal cancer and/or endometrial cancer[20]. Three of the *MSH6* variants classified as rare variants, including two synonymous exonic variants and one intronic variant. The prevalence of rare *MSH6* variants in their familial breast cancer cohort is similar to the prevalence we report here (3/38=7.9% versus 10.3% in our cohort) and supports our conclusion that rare *MSH6* variants are associated with familial breast cancer. Their and our observations both seem to point towards a currently unknown disease mechanism in breast carcinogenesis such as modulation of transcript expression levels or mediation of non-coding RNAs located in the genomic regions associated with these variants. Therefore, one may wonder whether this mechanism is similar to the as-yet unresolved disease mechanism underlying the more prevalent low-risk breast cancer alleles that recently have received much attention[21-26]

Table 3.2 Cancer incidence among 68 HBCC and 68 non-HBCC breast cancer families

	Number of cancer families		Number of cancers							
			BRC		CRC		Other cancers		Total cancers	
	HBCC	non-HBCC	HBCC	non-HBCC	HBCC	non-HBCC	HBCC	non-HBCC	HBCC	non-HBCC
Cancer families	–									
With three cancer cases	4	1	–	2	–	0	–	1	–	3
With four cancer cases	6	4	10	13	4	0	2	3	16	16
With five cancer cases	2	6	19	22	7	0	4	8	30	30
With six cancer cases	4	6	6	27	4	0	2	9	12	36
With seven cancer cases	11	9	17	41	6	1	5	21	28	63
With eight cancer cases	10	6	38	32	25	0	25	16	88	48
With nine cancer cases	31	8	37	45	18	2	35	25	90	72
With ten or more cancer cases	68	28	167	195	69	6	143	175	379	376
Total	540	68	294	377	133	9	216	258	643	644
Cancer cases	451	553								
With single tumors	78	473								
With double tumors	11	71								
With three or more tumors		9								

BRC breast cancer, CRC colorectal cancer, HBCC hereditary breast and colorectal cancer families; other cancer cases, cancer cases other than breast cancer or colorectal cancer

ACKNOWLEDGEMENTS

We wish to thank all members of the breast cancer families that have participated in this study. Financial support was provided by the Dutch Cancer Society (DDHK 2003-2862) and the Susan G. Komen Breast Cancer Foundation (BCT R0601309). MR is funded by the Higher Education Commission (HEC) of the Government of Pakistan.

REFERENCES

1. Meijers-Heijboer, H. et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
2. Vahteristo, P. et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**, 432-8 (2002).
3. Meijers-Heijboer, H. et al. The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet* **72**, 1308-14 (2003).
4. Wasielewski, M. et al. CHEK2 1100delC is a susceptibility allele for HNPCC-related colorectal cancer. *Clin Cancer Res* **14**, 4989-94 (2008).
5. Jass, J.R. Hereditary Non-Polyposis Colorectal Cancer: the rise and fall of a confusing term. *World J Gastroenterol* **12**, 4943-50 (2006).
6. Peltomaki, P. Lynch syndrome genes. *Fam Cancer* **4**, 227-32 (2005).
7. Vasen, H.F. et al. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J Med Genet* **44**, 353-62 (2007).
8. Bell, D.W. et al. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* **286**, 2528-31 (1999).
9. Lee, S.B. et al. Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. *Cancer Res* **61**, 8062-7 (2001).
10. Papadopoulos, N. et al. Mutations of GTBP in genetically unstable cells. *Science* **268**, 1915-7 (1995).
11. Vahteristo, P. et al. p53, CHK2, and CHK1 genes in Finnish families with Li-Fraumeni syndrome: further evidence of CHK2 in inherited cancer predisposition. *Cancer Res* **61**, 5718-22 (2001).
12. Bartek, J. & Lukas, J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**, 421-9 (2003).
13. Lee, J.S., Collins, K.M., Brown, A.L., Lee, C.H. & Chung, J.H. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature* **404**, 201-4 (2000).
14. Wang, Y. et al. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev* **14**, 927-39 (2000).
15. Hendriks, Y.M. et al. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. *Gastroenterology* **127**, 17-25 (2004).

16. Jensen, U.B. et al. Mismatch repair defective breast cancer in the hereditary nonpolyposis colorectal cancer syndrome. *Breast Cancer Res Treat* **120**, 777-82 (2010).
17. Wagner, A. et al. Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. *J Med Genet* **38**, 318-22 (2001).
18. Ramsoekh, D. et al. A high incidence of MSH6 mutations in Amsterdam criteria II-negative families tested in a diagnostic setting. *Gut* **57**, 1539-44 (2008).
19. Wasielewski, M., Elstrodt, F., Klijn, J.G., Berns, E.M. & Schutte, M. Thirteen new p53 gene mutants identified among 41 human breast cancer cell lines. *Breast Cancer Res Treat* **99**, 97-101 (2006).
20. Vahteristo, P. et al. No MSH6 germline mutations in breast cancer families with colorectal and/or endometrial cancer. *J Med Genet* **42**, e22 (2005).
21. Ahmed, S. et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-90 (2009).
22. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
23. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
24. Stacey, S.N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-9 (2007).
25. Stacey, S.N. et al. Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat Genet* **40**, 1313-8 (2008).
26. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-8 (2009).

Chapter 4

Low-risk susceptibility alleles in 40 human breast cancer cell lines

Muhammad Riaz, Fons Elstrodt, Antoinette Hollestelle, Abbas Dehghan, Jan G.M. Klijn and Mieke Schutte.
BMC Cancer 2009; 9: 236.

ABSTRACT

Introduction Low-risk breast cancer susceptibility alleles or SNPs confer only modest breast cancer risks ranging from just over 1.0 to1.3 fold. Yet, they are common among most populations and therefore are involved in the development of essentially all breast cancers. The mechanism by which the low-risk SNPs confer breast cancer risks is currently unclear. The breast cancer association consortium BCAC has hypothesized that the low-risk SNPs modulate expression levels of nearby located genes.

Methods Genotypes of five low-risk SNPs were determined for 40 human breast cancer cell lines, by direct sequencing of PCR-amplified genomic templates. We have analyzed expression of the four genes that are located nearby the low-risk SNPs, by using real-time RT-PCR and Human Exon microarrays.

Results The SNP genotypes and additional phenotypic data on the breast cancer cell lines are presented. We did not detect any effect of the SNP genotypes on expression levels of the nearby located genes *MAP3K1, FGFR2, TNRC9* and *LSP1*.

Conclusions The SNP genotypes provide a base line for functional studies in a well-characterized cohort of 40 human breast cancer cell lines. Our expression analyses suggest that a putative disease mechanism through gene expression modulation is not operative in breast cancer cell lines.

INTRODUCTION

About ten percent of breast cancer patients have a history of multiple breast cancer cases in their family, suggesting the inheritance of breast cancer susceptibility alleles in these families. Germline mutations in the *BRCA1* and *BRCA2* genes are identified in about one quarter of the families with breast cancer. Female carriers of *BRCA1* and *BRCA2* mutations have an estimated 50-90% life-time risk to develop breast cancer, classifying both genes as high-risk susceptibility genes[1-2]. Other high-risk breast cancer genes include the *TP53*, *PTEN* and *STK11* genes, but mutations in these genes account for only few familial breast cancers. *CHEK2* was the first moderate-risk breast cancer gene being identified[3-5]. Germline mutations in *CHEK2* are identified in upto 5% of breast cancer families, albeit that their prevalence varies widely among populations. Female carriers of *CHEK2* mutations have a moderate two to three fold increased risk to develop breast cancer. By now, several other moderate-risk breast cancer genes have been identified, including *ATM*, *BRIP1* and *PALB2*[6-9]. Mutations in these genes all confer increased breast cancer risks of two to three fold and mutations in each of these genes are identified in about 1% of the familial breast cancers. Recently, the international breast cancer association consortium (BCAC) has conducted a large genome-wide association study and identified five single nucleotide polymorphisms (SNPs) that associated with breast cancer[10]. Four of these SNPs were within haplotype blocks that contained genes: SNP rs2981582 locates in intron 2 of the *FGFR2* gene at chromosome 10q; SNP rs889312 locates near *MAP3K1* at 5q; SNP rs3803662 locates between *TNRC9* and the LOC643714 gene at 16q; and SNP rs3817198 locates intronic in *LSP1* at 11p. SNP rs13281615 locates at 8q24 in a region without any annotated genes. Importantly, independent genome-wide association studies have associated other SNPs in *FGFR2* with breast cancer[11-12]. As *FGFR2* had already been implicated in breast cancer[13-20], the significance of the *FGFR2* SNPs as susceptibility alleles seemed evident. The *TNRC9* SNP had also been associated with breast cancer in another study[21]. Lastly, the 8q24 SNP was of particular interest because other SNPs at 8q24 had been associated with increased risks of prostate cancer and colorectal cancer [22-26]. BCAC estimated that each of the five identified SNPs associated with rather small increased breast cancer risks, ranging from just over 1.0 to 1.3 fold, classifying them as low-risk susceptibility alleles[10]. However, these low-risk SNPs are very common and their impact is therefore still substantial, together accounting for almost 5% of the familial breast cancers.

The mechanism by which the low-risk susceptibility alleles confer breast cancer risks was obscure[10]. In analogy with the high-risk and moderate-risk breast cancer genes, it had been anticipated that the identified SNPs associated with disease-causing alleles in the coding sequences of nearby located genes. However, extensive sequencing efforts have not identified such alleles in the SNP-associated haplotype blocks, suggesting that the SNPs themselves might be the disease-causing susceptibility alleles[10]. BCAC therefore proposed an alternative disease mechanism that involves expression modulation of genes located in the vicinity of the identified SNPs, thereby conferring low breast cancer risks. Here, we have evaluated expression modulation in a well-characterized

cohort of 40 human breast cancer cell lines, allowing us to specifically address whether this mechanism might operate in breast cancer cells.

METHODS

Breast cancer cell lines

The 40 human breast cancer cell lines used in this study are listed in Table 1 and have been described in detail elsewhere[27]. Microsatellite analysis with nearly 150 polymorphic markers had shown that all cell lines are unique and monoclonal[28].

Genotyping

Genotypes of five low-risk susceptibility alleles have been determined: rs889312 (A>C) near the *MAP3K1* gene; rs2981582 (C>T) in the *FGFR2* gene; rs3803662 (C>T) near the *TNRC9* gene; rs3817198 (T>C) in the *LSP1* gene and rs13281615 (A>G) that located in a gene desert at chromosome 8q24[10]. Genotyping was performed by direct sequencing of PCR-amplified genomic templates, using the BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI 3130xL Genetic Analyzer. Primer sequences are available upon request.

Allele frequencies of cases and controls reported by BCAC have been obtained by using their reported Odds Ratio data[10], and inferring allele frequencies by assuming that Odds Ratios reflect the ratio of minor allele carriers versus major allele carriers from the cases divided by the ratio of minor allele carriers versus major allele carriers from the controls.

Expression analysis

Transcript expression levels of four genes have been determined: *MAP3K1*, *FGFR2*, *TNRC9* and *LSP1*. Quantitative real-time PCR (qRT-PCR) was performed on cDNA templates that had been generated with oligo-dT and random hexamer primers from total RNA isolates, using *Power* SYBR Green PCR Master Mix (Applied Biosystems) and an ABI Prism 7700. Ct values were normalized according *HPRT* and *HMBS* house keeper Ct values. Transcript expression had also been determined by Human Exon 1.0 ST microarrays (Affymetrix), as described elsewhere[29] The exon array data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE16732.

Statistical analysis

Statistical analyses were performed with Statistical Package for the Social Sciences (SPSS) version 11.5, considering P-values of less than 0.05 significant. Fisher's exact test was used to determine association of the SNP genotypes with the breast cancer cell lines. The Kruskal Wallis test was used to compare gene expression levels among three SNP genotype groups (major homozygotes, heterozygotes, and minor homozygotes).

RESULTS AND DISCUSSION

Genotyping of low-risk susceptibility alleles in breast cancer cell lines

Genotypes of five low-risk susceptibility alleles[10] were determined in a cohort of 40 human breast cancer cell lines. For each SNP, frequencies of major homozygotes, heterozygotes and minor homozygotes are shown in Figure 4.1a and genotypes are detailed in Table 4.1. Frequencies of homozygote genotypes typically were higher than anticipated, likely related to allelic losses in the cell line samples (Figure 4.1a;[10]). For four SNPs (*8q24*, *MAP3K1*, *FGFR2* and *TNRC9*), the minor allele frequencies among the cell lines were higher than among the 21,860 BCAC breast cancer cases and 22,578 population controls (Figure 4.1b;[10]). Fisher's exact testing indicated that the minor allele frequencies among the cell lines were significantly higher than the BCAC population controls for two SNPs: *MAP3K1* and *TNRC9* (Figure 4.1b). In Table 4.1 and 4.2, we also included previously-determined phenotypic and genotypic data on the breast cancer cell lines, including data on molecular subtyping and allelotyping[30,28]. Together with the SNP genotypes, we provide a base line for functional studies in this cohort of breast cancer cell lines.

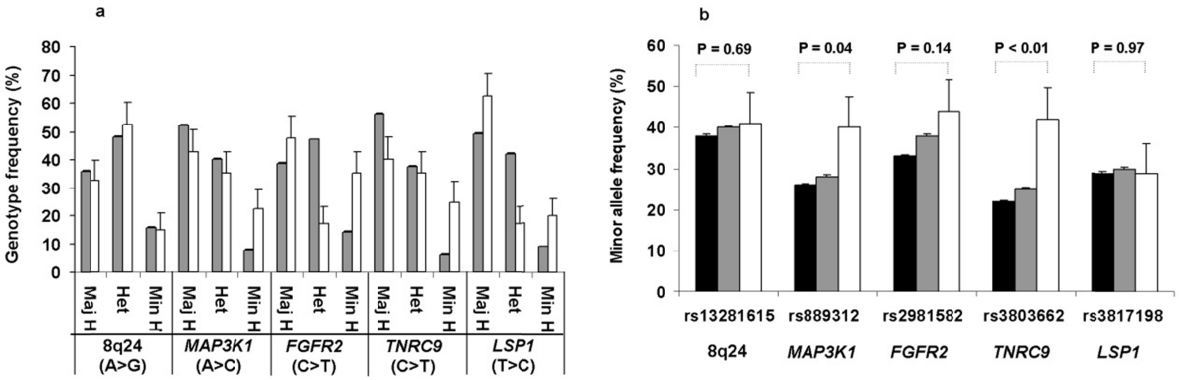


Figure 4.1 Genotypes and minor allele frequencies of five low-risk breast cancer susceptibility alleles or SNPs in human breast cancer cell lines. **4.1a.** Gray bars represent SNP genotype frequencies of 21,860 blood-derived samples from breast cancer cases reported by the breast cancer association consortium BCAC [10], and white bars represent genotype frequencies in 40 breast cancer cell lines. Maj H, major homozygotes; Min H, minor homozygotes; Het, heterozygote allele carriers. The major and minor alleles of each allele are indicated between brackets. **4.1b.** Black and gray bars represent minor allele frequencies in 22,578 population controls and 21,860 breast cancer cases, respectively, as reported by BCAC[10]. White bars represent frequencies identified in 40 breast cancer cell lines.

Table 4.1 Genotypes of five low-risk SNPs in 40 human breast cancer cell lines

Breast cancer cell lines	SNP genotypes and allelic losses									
	8q24	Loss 8q	*MAP3K1*	Loss 5q	*FGFR2*	Loss 10q	*TNRC9*	Loss 16q	*LSP1*	Loss 11p
SUM185PE	Het	No	Het	No	Maj H	nd	Min H	nd	Maj H	nd
BT483	Het	No	Maj H	No	Min H	Yes	Maj H	No	Maj H	No
MDA-MB-134VI	Het	No	Het	No	Het	No	Het	No	Maj H	No
MDA-MB-175VII	Het	No	Min H	No	Het	No	Het	No	Maj H	No
MDA-MB-415	Het	No	Het	No	Min H	Yes	Het	No	Het	No
MPE600	Het	No	Maj H	nd	Het	No	Maj H	nd	Het	No
SUM52PE	Maj H	nd	Maj H	nd	Maj H	nd	Maj H	nd	Maj H	nd
CAMA-1	Het	No	Het	No	Min H	No	Maj H	Yes	Maj H	Yes
MCF-7	Maj H	No	Het	No	Maj H	No	Het	No	Maj H	Yes
ZR75-1	Het	No	Het	No	Min H	Yes	Maj H	Yes	Het	No

Table 4.1 Continued

Breast cancer cell lines	SNP genotypes and allelic losses									
	8q24	**Loss 8q**	***MAP3K1***	**Loss 5q**	***FGFR2***	**Loss 10q**	***TNRC9***	**Loss 16q**	***LSP1***	**Loss 11p**
SUM44PE	Het	No	Min H	nd	Maj H	nd	Het	No	Min H	nd
T47D	Het	No	Het	No	Maj H	No	Min H	Yes	Min H	Yes
MDA-MB-361	Het	No	Het	No	Maj H	No	Het	No	Min H	No
BT474	Het	No	Maj H	No	Maj H	Yes	Maj H	No	Maj H	No
UACC812	Min H	No	Min H	No	Het	No	Het	No	Min H	Yes
ZR75-30	Maj H	nd	Min H	nd	Min H	nd	Min H	nd	Maj H	nd
OCUB-F	Maj H	nd	Maj H	nd	Min H	nd	Min H	nd	Maj H	nd
SK-BR-5	Het	No	Maj H	nd	Min H	nd	Min H	nd	Maj H	nd
SUM190PT	Min H	nd	Min H	nd	Maj H	nd	Min H	nd	Maj H	nd
SUM225CWN	Het	No	Maj H	nd	Maj H	nd	Maj H	nd	Het	No
MDA-MB-330	Het	No	Min H	Yes	Min H	Yes	Het	No	Maj H	No
MDA-MB-453	Het	No	Het	No	Min H	Yes	Maj H	Yes	Maj H	Yes
SK-BR-3	Min H	Yes	Het	No	Het	No	Maj H	Yes	Min H	No
EVSA-T	Min H	nd	Min H	nd	Maj H	nd	Maj H	nd	Min H	nd
UACC893	Maj H	No	Maj H	No	Maj H	Yes	Het	No	Maj H	No
BT20	Min H	No	Maj H	Yes	Maj H	Yes	Maj H	Yes	Het	No
HCC1937	Het	No	Min H	nd	Het	No	Het	No	Maj H	nd
MDA-MB-468	Maj H	Yes	Maj H	Yes	Maj H	Yes	Maj H	Yes	Maj H	Yes
SUM149PT	Min H	nd	Het	No	Min H	nd	Min H	nd	Maj H	Yes
SUM229PE	Maj H	nd	Maj H	nd	Maj H	nd	Maj H	nd	Min H	nd
BT549	Maj H	No	Maj H	No	Maj H	Yes	Het	No	Min H	Yes
Hs578T	Het	No	Maj H	Yes	Min H	Yes	Min H	No	Maj H	No
MDA-MB-157	Maj H	No	Maj H	Yes	Maj H	Yes	Min H	No	Maj H	Yes
MDA-MB-231	Maj H	Yes	Het	No	Het	No	Maj H	Yes	Het	No
MDA-MB-436	Maj H	Yes	Maj H	Yes	Maj H	No	Min H	No	Maj H	Yes
SK-BR-7	Het	No	Het	No	Min H	nd	Maj H	nd	Het	No
SUM159PT	Maj H	nd	Het	No	Min H	nd	Het	No	Maj H	nd
SUM1315MO2	Het	No	Maj H	nd	Min H	nd	Het	No	Maj H	nd
SUM102PT	Het	No	Het	No	Maj H	nd	Het	No	Maj H	nd
MDA-MB-435s	Maj H	No	Maj H	Yes	Maj H	Yes	Maj H	Yes	Maj H	No
Total major homozygotes	13		17		19		16		25	
Total heterozygotes	21		15		7		14		7	
Total minor homozygotes	6		8		14		10		8	
Percentage of allelic loss		13		25		52		32		37

Genotypes of five low-risk SNPs have been determined in the current study. Allelotype data have been reported eleswhere and involved microsatellite analysis. Loss: Yes, allelic loss at the indicated chromosomal region; and No, no allelic loss at the indicated chromosomal region. nd, not determined; 3-neg, triple-negative; Maj H, major homozygotes; Min H, minor homozygotes; Het, heterozygote allele carriers.

Expression levels of nearby located genes in breast cancer cell lines do not correlate with their SNP genotype

Surprisingly, BCAC has not identified disease-causing gene variants within the haplotype blocks of the five low-risk SNPs[10]. They proposed an alternative disease mechanism, in which SNP genotypes modulate expression levels of nearby located genes. Such disease mechanism was conceivable because the minor SNP alleles confer only low risks for breast cancer. Here, we have evaluated whether gene expression modulation is operative in breast cancer cell lines, by associating SNP genotypes of the breast cancer cell lines with the expression levels of nearby located genes.

Gene expression data of the four genes physically nearest to the SNPs were obtained by qRT-PCR analysis and by Affymetrix Human Exon 1.0 ST microarray profiling. Both transcript expression analysis methods revealed highly similar expression levels for each of the four genes: *MAP3K1*, *FGFR2*, *TNRC9* and *LSP1* with Spearman correlation coefficients of -0.6, -0.7, -0.8 and -0.4, respectively, among the 40 breast cancer cell lines (Table 4.3 and Figures 4.2 and 4.3).

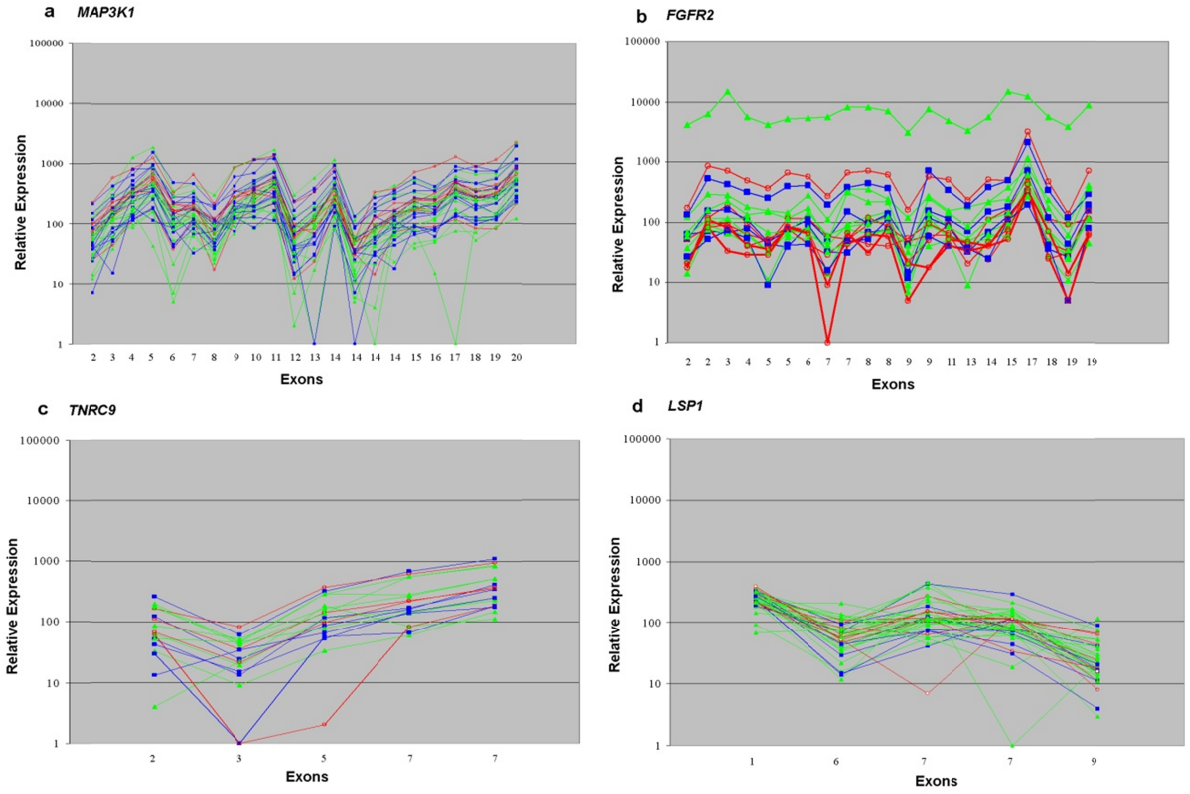


Figure 4.2 Normalized expression levels from Affymetrix Human Exon 1.0 ST microarrays of **4.2a**. *MAP3K1*; **4.2b**. *FGFR2*; **4.2c**. *TNRC9*; and **4.2d**. *LSP1*; are depicted from cell lines that had expression of the genes. Probe sets were ordered by physical location and indicated by exon, where probe sets that were not unique were deleted. Probe sets with expression values less than 50 were also deleted, unless more than 3 cell lines had expression levels higher than 100.

Because BCAC had shown that the low-risk SNPs confer breast cancer risks in a dose-dependent manner, with the highest risks for the minor homozygotes[10], association between gene expression levels and SNP genotypes was performed by three group comparisons. Exon array data are shown in Figure 4.2, with cell lines from each genotype group depicted in a different color. Unique outliers typically represented decreased expression of one or more probes sets, such as exon 17 of *MAP3K1* or exons 3– 5 of *TNRC9*, possibly related to the presence of SNPs in probe sequences, alternative splicing or genomic deletions[29]. Expression of recurrent isoforms (as reported by NCBI) was detected only for the *FGFR2* gene, with two cell lines expressing the isoform that lacked exon 9 (Figure 4.2). Both cell lines were minor homozygotes for the *FGFR2* SNP. Overall, there was no apparent association between the exon array expression level of each of the four genes and their SNP genotypes (Figure 4.2). The qPCR Ct-values are detailed in Table 4.3 and the three-group comparisons are shown in Figure 4.3. Again, we did not detect any association between gene expression levels with SNP genotypes for the four genes. It is possible that gene expression levels

are affected by allelic loss of the gene loci. We therefore also have compared gene expression levels in major and minor homozygotes with allelic loss to the gene expression levels in cell lines without allelic loss, but gene expression levels did not correlate with allelic losses either (Table 4.2).

Table 4.2 Molecular and phenotypic characterizations of 40 breast cancer cell lines

Breast cancer cell lines	Breast cancer subtype	Intrinsic subtype	Protein expression			
			ER	PgR	ERBB2	3-neg
SUM185PE	Luminal-type	Luminal	-	-	-	+
BT483	Luminal-type	Luminal	+	-	+	-
MDA-MB-134VI	Luminal-type	Luminal	+	-	-	-
MDA-MB-175VII	Luminal-type	Luminal	+	-	-	-
MDA-MB-415	Luminal-type	Luminal	+	-	-	-
MPE600	Luminal-type	Luminal	+	-	+	-
SUM52PE	Luminal-type	Luminal	+	-	+	-
CAMA-1	Luminal-type	Luminal	+	+	+	-
MCF-7	Luminal-type	Luminal	+	+	-	-
ZR75-1	Luminal-type	Luminal	+	+	+	-
SUM44PE	Luminal-type	Luminal	+	+	-	-
T47D	Luminal-type	Luminal	+	+	-	-
MDA-MB-361	Luminal-type	Luminal	+	+	++	-
BT474	Luminal-type	Luminal	-	+	++	-
UACC812	Luminal-type	Luminal	-	+	++	-
ZR75-30	Luminal-type	Luminal	+	-	++	-
OCUB-F	Luminal-type	Luminal	-	-	++	-
SK-BR-5	Luminal-type	Luminal	-	-	++	-
SUM190PT	Luminal-type	nd	-	-	++	-
SUM225CWN	Luminal-type	nd	-	-	++	-
MDA-MB-330	Luminal-type	ERBB2	+	-	++	-
MDA-MB-453	Luminal-type	ERBB2	-	-	++	-
SK-BR-3	Luminal-type	ERBB2	-	-	++	-
EVSA-T	Luminal-type	ERBB2	-	-	++	-
UACC893	Luminal-type	ERBB2	-	-	++	-
BT20	Basal-type	Basal-like	-	-	-	+
HCC1937	Basal-type	Basal-like	-	-	-	+
MDA-MB-468	Basal-type	Basal-like	-	-	-	+
SUM149PT	Basal-type	Basal-like	-	-	-	+
SUM229PE	Basal-type	Basal-like	-	-	-	+
BT549	Basal-type	Normal-like	-	-	-	+
Hs578T	Basal-type	Normal-like	-	-	-	+
MDA-MB-157	Basal-type	Normal-like	-	-	-	+
MDA-MB-231	Basal-type	Normal-like	-	-	-	+
MDA-MB-436	Basal-type	Normal-like	-	-	-	+
SK-BR-7	Basal-type	Normal-like	-	-	-	+
SUM159PT	Basal-type	Normal-like	-	-	-	+
SUM1315MO2	Basal-type	Normal-like	-	-	-	+
SUM102PT	Basal-type	Normal-like	nd	nd	nd	nd
MDA-MB-435s	Basal-type	Normal-like	-	-	-	+
Total phenotype positives			14	8	13	16

Phenotypic characterizations have been reported elsewhere and involved protein expression patterns of the cell lines for the breast cancer subtyping (cytokeratins, ER, PgR and ERBB2) and expression of the intrinsic gene set for the intrinsic subtyping[30]. Protein expression: +, expressed; ++, over expressed; -, not detectable; nd, not determined and 3-neg, triple-negative;

Altogether, these results strongly suggest that a putative disease mechanism by expression modulation does not operate via cancer cells. Yet, recent studies have shown that expression levels of the *FGFR2*, *MAP3K1* and *TNRC9* genes associated with their SNP genotype in clinical breast cancer samples[31-32]. It may be that expression modulation is operative in non-neoplastic stromal or epithelial cells and perhaps only early in carcinogenesis. Alternatively, it may be that expression modulation of these genes was operative in invasive breast cancer cells but was lost upon *in vitro* propagation of the cell lines. Expression analysis of carefully dissected tumor cells and non-neoplastic epithelial and stromal cells from clinical breast cancer samples should resolve this issue and may determine the precise mechanism of expression modulation by low-risk breast cancer susceptibility alleles.

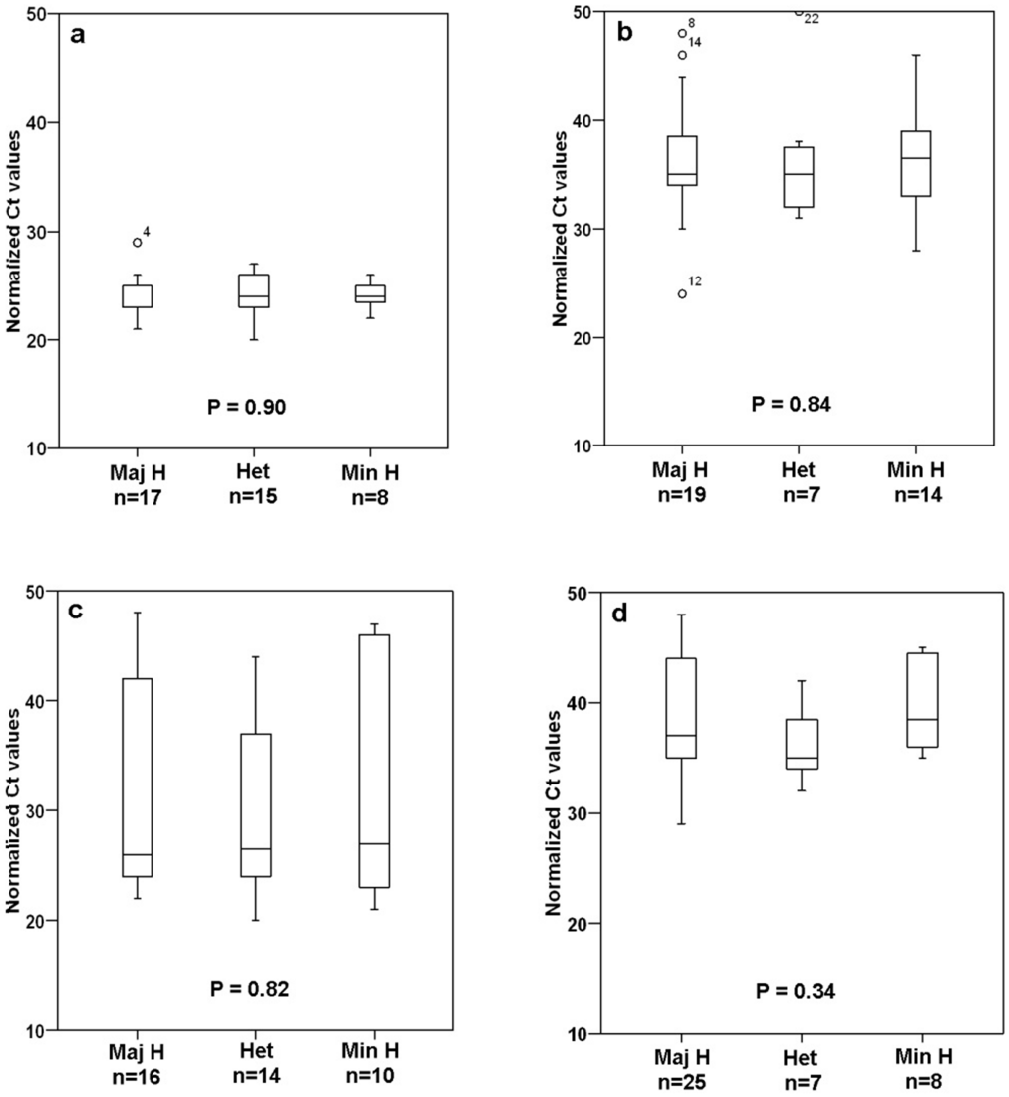


Figure 4.3 Correlation of gene expression levels of 4.3a. MAP3K1; 4.3b. FGFR2; 4.3c. TNRC9; and 4.3d. LSP1; with the genotypes of low-risk susceptibility alleles in 40 human breast cancer cell lines. Kruskal Wallis testing did not reveal any significant association between gene expression and SNP genotypes. Maj H, major homozygotes; Min H, minor homozygotes; Het, heterozygote allele carriers. The number of cell lines in each genotype group is indicated under the genotypes and data are detailed in Table 4.1.

Table 4.3 Gene expression analysis of *MAP3K1, FGFR2, TNRC9* and *LSP1* in 40 human breast cancer cell lines by quantitative RT-PCR represented by normalized Ct values

Breast cancer cell lines	Transcript expression (normalized Ct values)			
	MAP3K1	***FGFR2***	***TNRC9***	***LSP1***
BT20	23	30	45	38
BT474	24	35	22	35
BT483	22	31	25	36
BT549	29	32	39	40
CAMA-1	24	33	26	45
EVSA-T	25	33	25	45
HCC1937	26	32	37	45
Hs578T	26	45	45	34
MCF-7	25	34	28	39
MDA-MB-134VI	22	37	24	35
MDA-MB-157	25	44	33	29
MDA-MB-175VII	24	35	23	36
MDA-MB-231	26	45	45	35
MDA-MB-330	26	32	27	33
MDA-MB-361	23	34	23	44
MDA-MB-415	23	28	20	33
MDA-MB-435s	25	45	45	45
MDA-MB-436	26	35	31	37
MDA-MB-453	24	36	29	45
MDA-MB-468	25	35	37	37
MPE600	23	31	22	42
OCUB-F	23	39	23	42
SK-BR-3	25	32	26	37
SK-BR-5	23	37	21	43
SK-BR-7	26	39	45	35
SUM102PT	25	35	32	29
SUM1315M02	26	43	44	35
SUM149PT	27	38	45	37
SUM159PT	27	45	43	34
SUM185PE	23	38	23	45
SUM190PT	24	45	23	45
SUM225CWN	24	36	23	39
SUM229PE	25	39	33	36
SUM44PE	22	41	26	36
SUM52PE	24	24	24	37
T47D	20	36	45	35
UACC812	24	38	24	45
UACC893	21	36	25	36
ZR75-1	24	36	24	32
ZR75-30	23	33	23	43
Total high expressors (Ct≤20)	1	0	1	0
Total moderate expressors (Ct 20-30)	39	2	23	2
Total low expressors (Ct ≥30-35)	0	12	4	5
Total no expressors (Ct ≥35)	0	26	12	33

Table 4.4 Gene expression of MAP3K1, FGFR2, TNRC9 and LSP1 in human breast cancer cell lines according to their allelic loss status at the gene locus

Gene	Genotype (Number of cell lines)	Average expression level (Normalized Ct values)
MAP3K1	Het (n=14)	24±2
	Maj H no loss (n=4)	24±4
	Min H no loss (n=2)	24±0
	Maj H allelic loss (n=1)	26
	Min H allelic loss (n=6)	25±1
FGFR2	Het (n=7)	36±5
	Maj H no loss (n=4)	31±7
	Min H no loss (n=1)	33
	Maj H allelic loss (n=6)	35±6
	Min H allelic loss (n=6)	38±6
TNRC9	Het (n=14)	30±8
	Maj H no loss (n=2)	24±2
	Min H no loss (n=3)	36±8
	Maj H allelic loss (n=1)	45
	Min H allelic loss (n=7)	36±9
LSP1	Het (n=7)	37±4
	Maj H no loss (n=8)	37±4
	Min H no loss (n=2)	41±5
	Maj H allelic loss (n=3)	40±5
	Min H allelic loss (n=7)	38±6

Although numbers are small for some sample groups, there are no apparent differences in gene expression levels related to allelic loss status. Allelic loss data and qPCR expression data are detailed in Table 4.1 and Table 4.3, respectively. Maj H, major homozygotes; Min H, minor homozygotes; Het, heterozygote allele carriers.

CONCLUSION

We present the genotypes of five low-risk susceptibility alleles or SNPs of 40 human breast cancer cell lines. Using this cell line model, we have evaluated the BCAC hypothesis that low-risk SNPs confer breast cancer risks by modulation of expression levels of nearby located genes. We found no evidence for expression modulation in the breast cancer cell lines, suggesting that such disease mechanism is more likely to operate in non-neoplastic epithelial or stromal cells or has been lost during *in vitro* propagation of the cell lines.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

MR and FE carried out genotyping and transcript expression analyses, AH carried out protein expression analyses, and AD performed part of the statistical analyses. MR, JGMK and MS designed

the study, and MR and MS wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

MR is funded by the Higher Education Commission (HEC) of the Government of Pakistan. Financial support was provided by grants from the Susan G. Komen Breast Cancer Foundation (BCT R0601309) and Erasmus MC Mrace 2007.

REFERENCES

1. Thompson, D. & Easton, D. The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia* **9**, 221-36 (2004).
2. Walsh, T. & King, M.C. Ten genes for inherited breast cancer. *Cancer Cell* **11**, 103-5 (2007).
3. Meijers-Heijboer, H. et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
4. Vahteristo, P. et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**, 432-8 (2002).
5. Consortium, C.B.C.C.-C. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* **74**, 1175-82 (2004).
6. Renwick, A. et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-5 (2006).
7. Seal, S. et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-41 (2006).
8. Rahman, N. et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**, 165-7 (2007).
9. Stratton, M.R. & Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**, 17-22 (2008).
10. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
11. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
12. Gold, B. et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A* **105**, 4340-5 (2008).
13. Adnane, J. et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* **6**, 659-63 (1991).
14. Luqmani, Y.A., Graham, M. & Coombes, R.C. Expression of basic fibroblast growth factor, FGFR1 and FGFR2 in normal and malignant human breast, and comparison with other normal tissues. *Br J Cancer* **66**, 273-80 (1992).

15. Penault-Llorca, F. et al. Expression of FGF and FGF receptor genes in human breast cancer. *Int J Cancer* **61**, 170-6 (1995).
16. Dickson, C., Spencer-Dene, B., Dillon, C. & Fantl, V. Tyrosine kinase signalling in breast cancer: fibroblast growth factors and their receptors. *Breast Cancer Res* **2**, 191-6 (2000).
17. Heiskanen, M. et al. CGH, cDNA and tissue microarray analyses implicate FGFR2 amplification in a small subset of breast tumors. *Anal Cell Pathol* **22**, 229-34 (2001).
18. Antoniou, A. et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**, 1117-30 (2003).
19. Moffa, A.B., Tannheimer, S.L. & Ethier, S.P. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol Cancer Res* **2**, 643-52 (2004).
20. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).
21. Stacey, S.N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-9 (2007).
22. Gudmundsson, J. et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* **39**, 631-7 (2007).
23. Haiman, C.A. et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* **39**, 638-44 (2007).
24. Haiman, C.A. et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* **39**, 954-6 (2007).
25. Tomlinson, I. et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* **39**, 984-8 (2007).
26. Zanke, B.W. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* **39**, 989-94 (2007).
27. Elstrodt, F. et al. BRCA1 mutation analysis of 41 human breast cancer cell lines reveals three new deleterious mutants. *Cancer Res* **66**, 41-5 (2006).
28. Harkes, I.C. et al. Allelotype of 28 human breast cancer cell lines and xenografts. *Br J Cancer* **89**, 2289-92 (2003).
29. Schutte, M. et al. Exon expression arrays as a tool to identify new cancer genes. *PLoS ONE* **3**, e3007 (2008).
30. Hollestelle, A. et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* **121**, 53-64 (2010).
31. Nordgard, S.H. et al. Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. *Breast Cancer Res* **9**, 113 (2007).
32. Meyer, K.B. et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* **6**, e108 (2008).

Chapter 5

Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival and mRNA expression of the nearest genes

Muhammad Riaz, Els MJJ Berns, Anieta M Sieuwerts, Kirsten Ruigrok-Ritstier, Vanja de Weerd, Arwin Groenewoud, Andrè G Uitterlinden, Maxime P Look, Jan GM Klijn, Stefan Sleijfer, John A Foekens and John WM Martens.

Breast Cancer Res Treat. 2012; 133(3):843-51

ABSTRACT

To understand the biology of low-risk breast cancer alleles and to investigate whether these loci also contribute to disease progression once established, we examined the association of SNPs tagging the low-risk breast cancer loci in or near *FGFR2*, *LSP1*, *MAP3K1, H19*, *TOX3*, *POU5F1P1*, *MYC* and 2q35, with clinical, pathological characteristics, prognosis, and mRNA expression of the nearest genes. Tumor DNA samples of 2,480 breast cancer patients were available. Out of this cohort, in 1,290 patients with lymph-node negative disease who did not receive adjuvant systemic therapy, SNP status was associated with metastasis-free survival (MFS). In 1,401 patients, the mRNA expression levels of *FGFR2*, *LSP1*, *MAP3K1, H19*, *TOX3*, *POU5F1P1* and *MYC* were determined and correlated with SNP genotypes. The SNP rs2981582 in *FGFR2* was significantly associated with positive ER and PgR status (P <0.001 and P =0.003, respectively). No other significant associations with patient or tumor characteristics were observed. Only rs2107425 near *H19* was significantly associated with shorter MFS in uni- and multivariate analysis (HR: 1.53, CI: 1.12-2.08, P = 0.006 and HR: 1.59, CI: 1.16-2.20, P =0.004, respectively), with the more aggressive minor allele displaying a recessive trait. The minor allele of SNP rs3803662 located near the *TOX3* gene was associated with lower mRNA expression of this gene. In conclusion, except for the association of rs13283662 with *TOX3* gene expression indicating a tumor suppressor role of *TOX3*, our findings suggest that breast cancer low-risk loci generally do not affect expression of the nearest gene in breast tumor tissue. Also the prognosis of patients is largely not affected by low-risk breast cancer loci except for the SNP near *H19*. How, this SNP affects prognosis warrants further study as it does not operate through altering *H19* mRNA expression.

INTRODUCTION

Genome-wide association studies (GWAS) have recently identified various single nucleotide polymorphisms (SNPs) tagging several breast cancer predisposing loci[1-3]. The identified SNPs are: rs2981282 located within intron 2 of the fibroblast growth factor receptor 2 (*FGFR2)*; rs3817198 in the lymphocyte-specific protein 1 (*LSP1*) gene; rs889312 located close to the mitogen-activated protein kinase kinase kinase 1 (*MAP3K1)* gene; rs3803662, rs12443621 and rs8051542 located near the trinucleotide repeat containing 9 (*TNRC9*) gene, also known as *TOX3*; rs2107425 located close to the imprinted *H19* gene and rs13387042 and rs13281615 in the 2q35 and 8q24 regions, respectively, which lack nearby annotated genes. The most nearest genes to rs13281615 in the 8q24 region are *POU5F1P1*, a transcription factor with a POU homeodomain, and *MYC*, a proto-oncogene. Although several independent GWAS have replicated the association of these SNPs with increased cancer risk[3-6], the mechanisms by which these loci in the target tissue may exert their effect are poorly understood.

In analogy to high-risk and moderate-risk breast cancer genes, it was speculated that the identified SNPs tag causal variants in the coding region of nearby genes[1]. However, extensive sequencing efforts of the nearest gene by[1] have not identified such variants in the SNP-associated haplotype blocks. This suggests that an alternative disease mechanism in which low-risk loci might involve expression modulation of nearby located genes and hence predispose to breast cancer. For example, Meyer *et al* observed an association of risk alleles in *FGFR2* with increased *FGFR2* mRNA expression in 63 invasive breast cancer samples [7], and Sun *et al* reported an association of the normal allele with higher *FGFR2* mRNA expression in 81 normal breast tissue samples [8]. Besides that these results appear contradictory, both studies lacked substantial power to test, conclusively, the hypothesis of expression modulation of the nearest genes by risk loci.

To further test the hypothesis that in breast cancer these SNPs may operate through expression modulation of the nearest genes, we measured in a large subset of 1,401 primary breast cancer samples, mRNA transcript levels of the most plausible genes located nearest to the respective SNPs. In addition, we investigated whether these low-risk breast cancer loci not only relate to onset of breast cancer, as previously reported, but also contribute to its progression once established. Therefore, we have studied in the present study the association of SNP genotypes of seven low-risk breast cancer loci with patient and clinico-pathological tumor characteristics, as well as with tumor aggressiveness.

MATERIALS AND METHODS

Patients and tumor characteristics

This retrospective study was approved by the medical ethical committee of the Erasmus Medical Centre Rotterdam, the Netherlands (MEC 02.953). The study was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.federa.org/?s=1&m=99), consent was not required, and, wherever possible, has been

reported in line with the Reporting Recommendations for Tumor Marker Prognostic Studies guidelines[9]. Frozen primary tumor samples from patients who entered the clinic between 1978 and 2004, and from whom detailed clinical follow-up was available, were used. The median follow up of patients alive was 106 (range, 3-315) months and mean age was 55.9 (range, 22-88) years.
Tumor estrogen receptor (ER) and progesterone receptor (PgR) analyses were performed by routine ligand binding assay or enzyme-linked immunoassay[10]. The cutoff used to classify the tumors as ER and/or PgR positive was 10 fmol/mg cytosolic protein. Patients with missing values for ER and PgR were indicated as an unknown category in Table 5.1. Details of patients and tumor characteristics are presented in Table 5.1. Transcript level expression of nearby located candidate genes and *HER2* were determined in a subset of 1,401 tumors. Details of patients and tumor characteristics of this subgroup are presented in Supplementary Table S5.1.

SNP genotyping

Genomic DNA of 680 tumor homogenates has been isolated as described previously[11] and the genomic DNA of 1,800 tissue sections were isolated from two to ten 30 µm cryostat sections (5 to 20 mg) with the NucleoSpin®Tissue kit (Macherey-Nagel; Bioké, Leiden, The Netherlands) according to the protocol provided by the manufacturer. The quantity and quality of the isolated DNA was established by ultraviolet spectroscopy, by examination of the product size after agarose gel electrophoresis, and by the ability of the sample to be linearly amplified by real-time PCR in a serial dilution with a set of primers located in an intron of the hydroxymethylbilane synthase on chromosome 11 and thymidine kinase on chromosome 17[12]. Due to low sensitivity of the assay, SNP rs13281615 in the 8q24 region was genotyped in whole genome amplified (WGA) DNA of 1,705 tumors as described before[12]. Briefly, 10 ng aliquots of genomic DNA of 1,705 tumors were amplified with the GenomiPhi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ, USA) according to the protocol provided by the manufacturer[12], typically yielding 4 µg amplified genomic DNA with the 20 kb band still visible on gel. Genotyping was performed by a fluorescent 5′ exonuclease assay (TaqMan) using the ABI PRISM 7900HT Sequence Detection Sequence (PE Biosystems) according to the manufacturer's instructions. Primers and probes for SNPs rs2981582 (C_2917302_10), rs3817198 (C_27493923_10), rs889312 (C_8886795_10), rs3803662 (C_25968567_10), rs12443621 (C_30765260_10), rs8051542 (C_1230104_10), rs13387042 (C_32048042_10), rs13281615 (C_1332250_10) were purchased from Applied Biosystems as Made to Order, and for rs2107425 (forward: ATA ATG CCC GAC CTG AAG ATC TG, reverse: GCG TCG CAG GGT TCA C , VIC-probe: CAC TCA TGG GAG CCG, FAM-probe: ACA CTC ATA GGA GCC G) as Assay on Demand. All reactions were carried out with no template as negative controls. Duplicate samples (n = 352) extracted with the two different DNA isolation procedures as mentioned above were analyzed to assess concordance and quality of genotyping. High concordance was found between the SNPs mentioned indicating that both procedures were equivalent. Due to sensitivity issue of the assays, the SNP genotypes of some patients could not be determined and were regarded as missing values for the corresponding SNPs. The missing values for each of the analyzed SNPs are also presented in Table 5.1.

Table 5.1 Associations of seven SNPs with patient and tumor characteristics (n = 2,480)

		SNPs	rs2981582			rs3817198			rs889312			rs2107425			rs3803662			rs13281615			rs13387042		
		Nearest genes	*FGFR2*			*LSP1*			*MAP3K1*			*H19*			*TOX3*			8q24			2q35		
Patients with genotype data (missing values)			n = 2,410 (70)			n = 2,445 (35)			n = 2,460 (20)			n = 2,430 (50)			n = 2,379 (101)			n = 1,705 (775)			n = 2,362 (118)		
		SNP genotypes	CC	CT	TT	TT	TC	CC	AA	AC	CC	GG	GA	AA	GG	GA	AA	AA	AG	GG	GG	GA	AA
Characteristics	**Subgroups**	**Patients**	33	49	18	49	42	9	48	43	9	48	43	9	51	40	9	35	47	18	28	48	23
		n **%**																					
Age (years)	<40	305 12	34	48	18	49	43	8	43	45	12	44	47	9	52	38	10	39	43	18	26	49	25
	41-55	971 39	34	48	18	49	42	9	49	42	9	51	41	8	49	41	10	34	48	18	30	48	23
	56-70	799 32	31	50	19	50	42	9	48	44	8	46	45	9	51	40	9	35	47	19	29	47	25
	>70	405 16	34	50	16	47	43	9	52	39	9	50	39	11	56	37	7	36	46	18	26	52	21
			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Menopausal status	Pre	1,078 43	34	49	18	48	42	9	47	43	10	50	42	8	49	41	9	35	48	17	28	49	24
	Post	1,402 57	33	49	18	49	42	9	50	42	8	48	43	9	53	39	9	35	46	19	29	48	23
			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor size	pT1	910 37	32	50	18	47	44	9	48	43	9	48	43	10	51	41	9	34	48	17	28	48	24
	pT2	1,293 52	34	48	18	50	41	9	49	43	8	49	43	8	51	40	10	36	45	19	29	48	24
	pT3	277 11	34	48	18	49	42	8	50	40	10	46	43	11	56	36	7	34	47	19	27	52	22
			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Lymph nodes involved	0	1290 52	33	49	18	49	42	8	48	42	9	50	42	8	51	39	10	35	46	19	28	47	24
	1-3	550 22	32	50	18	46	43	10	49	42	9	47	44	9	51	41	8	36	45	18	28	50	22
	>3	640 26	33	48	19	49	42	9	48	44	8	47	43	10	53	39	8	34	50	16	29	48	23
			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor grade	Poor	1,352 55	35	49	16	49	42	9	49	43	8	48	43	9	53	38	9	35	47	18	29	47	24
	Good/moderate	435 18	34	48	18	45	45	10	44	46	11	49	42	8	47	43	10	36	47	17	27	49	23
	Unknown	693 28	29	50	22	50	41	9	50	41	9	48	43	9	51	41	8	35	45	20	28	50	22
			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor histology	IDC	1,341 54	33	49	18	48	43	10	47	44	8	51	41	8	50	42	8	35	46	19	28	48	24
	ILC	190 8	26	54	20	47	45	8	52	39	9	48	42	10	51	40	9	39	43	18	39	46	15
	Muc	57 2	38	48	14	42	46	12	50	36	14	44	46	11	50	43	7	23	44	33	26	50	24
	Medul	57 2	34	55	11	60	33	7	54	38	9	44	55	2	48	44	7	32	44	24	30	49	21
	DCIS+IDC	315 13	32	50	18	51	41	8	48	41	11	44	44	12	51	38	11	32	54	14	28	46	26
	Unknown	520 21	35	46	19	50	42	8	49	42	9	47	45	9	56	33	11	38	45	17	27	51	22
			n.s.			n.s.			n.s.			n.s.			n.s			n.s.			n.s.		

Table 5.1 Continued

		SNPs		**rs2981582**			**rs3817198**			**rs889312**			**rs2107425**			**rs3803662**			**rs13281615**			**rs13387042**		
		Nearest genes		***FGFR2***			***LSP1***			***MAP3K1***			***H19***			***TOX3***			**8q24**			**2q35**		
Patients with genotype data (missing values)				n = 2,410 (70)			n = 2,445 (35)			n = 2,460 (20)			n = 2,430 (50)			n = 2,379 (101)			n = 1,705 (775)			n = 2,362 (118)		
		SNP genotypes		**CC**	**CT**	**TT**	**TT**	**TC**	**CC**	**AA**	**AC**	**CC**	**GG**	**GA**	**AA**	**GG**	**GA**	**AA**	**AA**	**AG**	**GG**	**GG**	**GA**	**AA**
Estrogen receptor	Negative	655	26	41	44	15	53	38	9	52	41	7	48	43	9	53	38	9	36	46	18	27	47	26
	Positive	1,809	73	30	51	19	47	44	9	47	43	10	49	43	9	51	40	9	35	47	18	29	49	22
	Unkown	16	1	44	44	13	25	56	19	50	50	0	56	44	0	53	40	7	44	33	22	38	25	38
				***P* < 0.001**			***P* = 0.03**			***P* = 0.03**			n.s.			n.s.			n.s.			n.s.		
Characteristics	**Subgroups**	**Patients**		33	49	18	49	42	9	48	43	9	48	43	9	51	40	9	35	47	18	28	48	23
		n	**%**																					
Progesterone receptor	Negative	782	32	38	46	16	52	39	9	50	43	7	49	41	10	52	41	8	36	44	20	28	46	26
	Positive	1,555	63	31	50	19	47	44	9	48	43	10	49	43	8	52	39	10	36	47	17	29	49	22
	Unknown	143	6	31	52	17	52	36	11	48	45	8	43	47	9	47	44	9	29	56	15	25	55	20
				***P* = 0.003**			***P* = 0.05**			n.s.			n.s.			n.s.			n.s.			n.s.		
[a]*HER2* status (mRNA)	Negative	1,146	82	34	49	18	50	41	9	49	42	9	48	43	9	50	41	9	35	46	19	27	49	24
	Positive	234	17	32	49	19	53	39	8	43	50	7	49	42	9	55	38	7	34	48	18	31	45	24
	Unknown	20	1	40	55	5	45	45	10	60	25	15	30	45	25	47	42	11	40	40	20	30	50	20
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		

All *P*-values were derived from Pearson's chi-squared statistics. Minor variations in percentages of the patients among SNPs occur due to variation in missing values per SNP. *n.s.*, non-significant; [a] *HER2* mRNA data is available only for 1,401 tumors.

RNA isolation and quantitative RT-PCR

RNA isolation, cDNA synthesis, quantification of specific mRNA species, and quality control checks were done as described in detail elsewhere[13]. In brief, total RNA was extracted with RNA-Bee (Campro, Veenendaal, the Netherlands) according to the manufacturer's instructions and 2 to 5 µg of total RNA sample aliquots were reverse transcribed with oligo(dT)12-18 and random hexamer primers using the RevertAid H Minus First Strand cDNA Synthesis Kit from Fermentas (St. Leon-Rot, Germany) or the Superscript II RNase H- kit from Invitrogen (Breda, the Netherlands), respectively, according the protocols provided by the manufacturers. Prior to PCR, the resulting cDNA samples were treated with RNase H- (Ambion, Huntingdon, United Kingdom). The quantity and quality of the isolated RNA was established by UV spectroscopy, by examination of rRNA bands after agarose gel electrophoresis, and by the ability of the sample to be linearly amplified in a serial dilution with our housekeeping gene set (see next section for further details). Samples of total RNA not showing both the 18S and 28S bands (6%) or at 15 µg reverse-transcribed total RNA not amplifiable within 26 cycles at our fixed threshold value of 0.02 with our housekeeping set were excluded. Real-time quantitative PCR was done in a Mx3000P Real-time PCR System (Agilent, Amsterdam, The Netherlands) according to the recommended protocol. Commercially available Taqman Gene Expression Assays (Applied Biosystems) were used for *FGFR2* (Hs00240792_m1), *MAP3K1* (Hs00394890_m1), *TOX3* (Hs00300355_m1), *LSP1* (Hs00158885_m1), *H19* (Hs00399294_g1), *POU5F1/POU5F1P3/POU5F1P1* (Hs00999634_gH) and *MYC* (Hs00905030_m1). Forty rounds of amplification were performed according to the supplier's protocol and at the end of the amplification fluorescent signals of the TaqMan probes were used to generate cycle threshold (Ct) values from which mRNA expression levels were calculated. To enable comparison of the levels of specific mRNAs in different samples, they were evaluated relative to the average expression levels of three reference genes, porphobilinogen deaminase *(PBGD)*, hypoxanthine-guanine phosphoribosyltransferase *(HPRT)*, and β-2-microglobulin *(β2M)*. Expression levels of the target genes relative to the average expression levels of three reference genes was quantified as follows: mRNA target = $2^{(\text{mean Ct reference-mean Ct target})}$. The primer sequences for *ER-α*, *PgR*, *ERBB2/HER2* and for the three reference genes, *PBGD*, *HPRT*, and *β2M*, have all been previously described[13-14].

Statistical analyses

Pearson's χ^2 analysis was used to test for independence of the alleles [Hardy-Weinberg equilibrium (HWE)], the allelic distribution in breast tumors, to compare genotype frequencies between the three groups, and to check the relation of SNPs with patients and tumor characteristics. The hazard ratios (HRs) and their 95% confidence interval (95% CI) for the SNP alleles and traditional prognostic factors were determined with Cox proportional hazards models for both univariate and multivariate metastasis-free survival (MFS) analyses. MFS was defined as a time point before which a patient was found to be free from metastasis to distant organs. Differences between HR for SNP carrier vs. non-carriers were tested using the likelihood ratio test associated with the Cox regression analysis. The assumption of proportional hazards was investigated using a test based on the Schoenfeld residuals and was not violated for SNP rs2107425. In the multivariate model, the

contribution of the SNP alleles was adjusted for the classical prognostic factors; age, menopausal status, nodal status, tumor size, differentiation grade, and receptor status. MFS probabilities were calculated using the actuarial method of Kaplan-Meier and a log-rank test was used to test for differences between the survival curves.

The association of SNPs and mRNA expression was determined using the Kruskal-Wallis test. A Bonferroni correction was applied to address the problem of multiple comparisons. All computations were carried out using the STATA statistical package, version 11.1 (Stata Corp., College Station, TX). A *P*-value ≤0.05 was considered statistically significant. All *p*-values are two-sided and relate to all data during the total period of follow-up.

RESULTS

Frequencies of low-risk alleles in tumor samples and association with patient and tumor characteristics

We genotyped nine SNPs tagging linkage disequilibrium (LD) blocks containing seven loci, which have been reported to be associated with increased breast cancer risk[1-2,15], in target tissue of breast cancer tumors from a cohort of 2,480 patients. Genotype determination and test for Hardy-Weinberg equilibrium (HWE) revealed that all nine genotyped SNPs in this cohort were in Hardy-Weinberg equilibrium ($P \geq 0.08$ for all SNPs). The calculated minor allele frequencies (MAF) of the SNPs in our cohort are given in Table 5.2. To check whether risk alleles showed an association with specific clinico-pathological characteristics of the patients and tumors, we investigated the correlation of the SNPs with patient and tumor characteristics (Table 5.1). The risk alleles of the SNP rs2981582 in *FGFR2* were significantly associated with ER- and PgR-positive tumors ($P < 0.001$ and $P = 0.003$, respectively). None of the other SNPs showed a strong association with either patient or tumor characteristics. After Bonferroni correction, only the association of *FGFR2* allele with ER and PgR level remained significant.

Table 5.2 Minor allele frequencies of low-risk breast cancer loci in the target tissues.

SNP	Locus	Nearest gene	MAF (European population)	MAF (Target tissue)
rs2881582	10q26	*FGFR2*	0.38	0.42
rs3817198	11p15	*LSP1*	0.30	0.30
rs889312	5q11	*MAP3K1*	0.28	0.30
rs2107425	11p15	*H19*	0.31	0.30
rs3803662	16q12	*TOX3*	0.25	0.29
rs12443621	16q12	*TOX3*	0.46	0.46
rs8051542	16q12	*TOX3*	0.44	0.45
rs13281615	8q24	*POU5F1P1/MYC*	0.40	0.58
rs13387042	2q35	Not analyzed	0.52	0.50

Association of low-risk alleles with prognosis of primary breast cancer

One of the aims of this study was to study a possible prognostic value of the low-risk SNPs. As prognosis is confounded by factors such as adjuvant treatment, we limited our analysis to our 1,290 patients with lymph node-negative disease who did not receive any adjuvant systemic

therapy. MFS was used as the primary end-point in this analysis. Compared with wild-type allele, patients carrying homozygous risk alleles AA of rs2107425 in *H19* were associated with a poor MFS in both univariate (HR: 1.53, CI: 1.12-2.08, *P* =0.006) and multivariate analysis (HR: 1.59, CI: 1.16-2.20, *P* =0.004) corrected for the base model that included age, menopausal status, pathological tumor size and grade, and ER and PgR protein levels (Supplementary Table S5.2). Kaplan-Meier survival curves were used to visualize the significant worse MFS (Logrank test *P*= 0.006) of patients homozygous for the risk allele of rs2107425 (Figure. 5.1). None of the other SNPs were related with the length of MFS in this cohort of 1,290 untreated lymph node-negative patients.

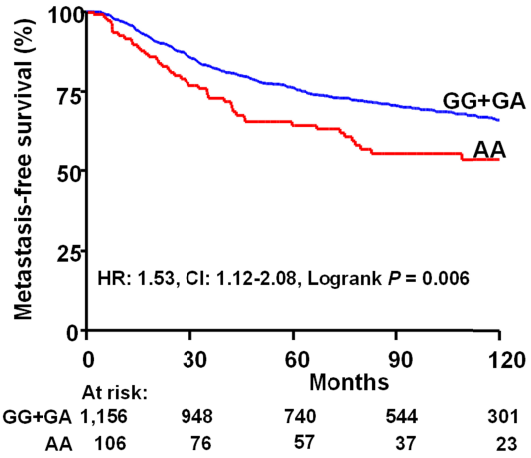


Figure 5.1 Kaplan-Meier analysis of metastasis-free survival for rs2107425 SNP near *H19* in lymph node-negative patients who did not receive any adjuvant systemic therapy (n = 1,290). The number of patients at risk on various time intervals is indicated.

Correlation of SNP genotype with expression changes of the nearest genes

The second aim of the study was to determine whether the risk alleles were associated with altered expression of the nearest gene. Therefore, we analyzed mRNA expression of the most plausible genes located close to the disease associated SNPs in a subset of 1,401 patients for which RNA was available (Figure 5.2; clinical characteristics of this cohort are provided in Supplementary Table S5.1). For three SNPs near the *TOX3* gene (rs3803662, rs12443621, rs8052542), which all have been associated with cancer risk, we found significant associations of SNPs rs3803662 and rs12443621 with a lower expression of *TOX3* mRNA (*P* =0.0019 and *P* =0.029, respectively; Kruskal-Wallis test). The lower *TOX3* mRNA expression was dose dependent and decreased with the number of risk alleles (*P* =0.02 for trend). However, rs8052542 near *TOX3*, which has the weakest association with breast cancer risk, did not show an association with *TOX3* mRNA expression. Correlation analysis of these three SNPs near *TOX3* indicated that the highest risk-associated SNP rs3803662 showed a stronger correlation with rs12443621 than with rs8051542 (Pearson correlation coefficient r = 0.65 and r = 0.20, respectively). For none of the other loci investigated we did observe a significant correlation of tumor mRNA expression of the analyzed nearby gene(s) with the tumor genotype of the corresponding SNP (Figure 5.2). This includes the risk conferring SNP rs2981582 previously connected to expression modulation of *FGFR2*[7]. Also the reported physical interactions in breast cancer for rs13281615 in 8q24 with the *MYC* gene[16] did not translate into an altered *MYC* mRNA

expression in primary tumors of our cohort (Figure 5.2). Furthermore, the genotype of this SNP in the subset of breast tumors for which amplification status was available (n = 163) did not correlate with *MYC* amplification (Supplementary Table S5.3).

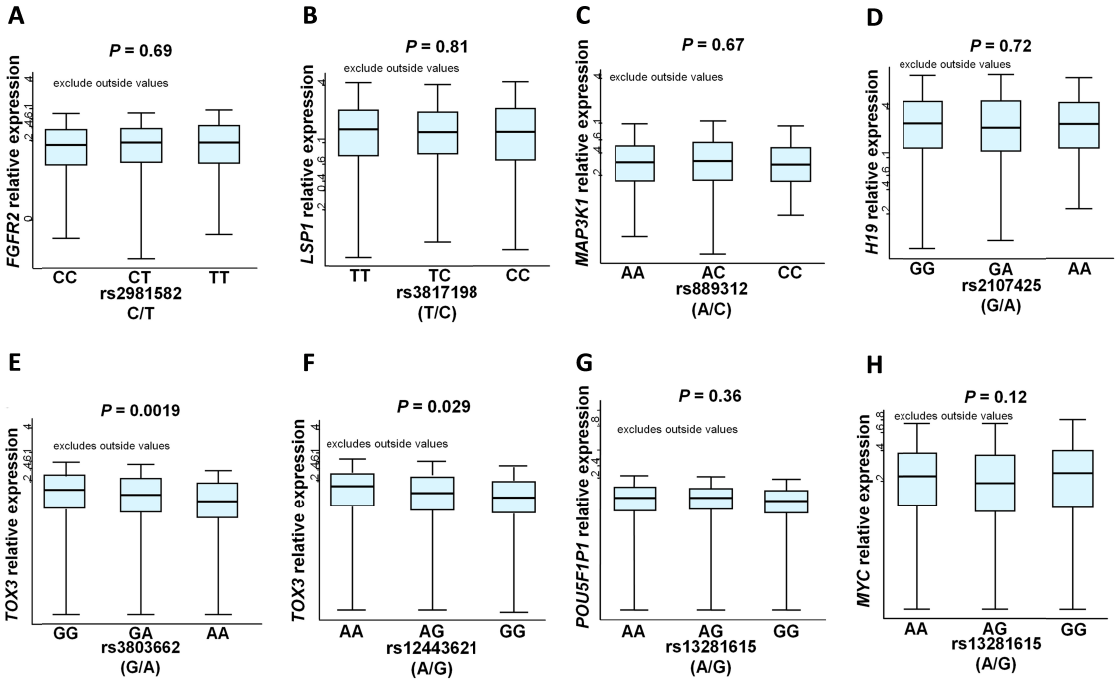


Figure 5.2 Correlation of mRNA expression level of genes with genotype of their respective SNPs in 1,401 primary breast tumors: **A.** *FGFR2*; **B.** *LSP1*; **C.** *MAP3K1*; **D.** *H19*; **E.** *TOX3*; **F.** *TOX3*; **G.** *POU5F1P1* and **H.** *MYC*. Kruskal-Wallis test was used to test for significant differences between the respective three genotype groups. Allele changes go from major allele to minor allele (major/minor). The rs3803662 and rs12443621 SNPs are located on the same LD block previously shown to be associated with breast cancer risk and these SNPs are correlated with each other (r = 0.65).

DISCUSSION

Several large case-controls studies have discovered various SNPs being associated with breast cancer risk[1,17-18]. Among these newly discovered SNPs, the associations of nine SNPs tagging seven low-risk breast cancer loci have been independently replicated in many populations of different ethnic backgrounds[3-6]. Following their discovery, some of these loci were found to be associated with important clinical and pathological characteristics of the resulting invasive breast cancer. Rs2981582 in *FGFR2* and rs13282615 in 8q24 have been shown to be related to ER-positive and PgR-positive tumors, whereas rs12383662 in *TOX3* was associated with ER-negative tumors[19]. Rs13387042 on 2q35 has been shown to be associated with both ER-positive and ER-negative breast cancer[20]. Consistent with previous findings, we also found rs2981582 in *FGFR2* to be associated with ER-positive and PgR-positive tumors in our cohort. The other associations with hormone receptor status could not be confirmed.

The fact that genetic changes in tumor tissues might affect SNP genotype analysis we determined copy number variation in a subset of 313 patients[21]. The association of copy number variation in the region containing risk-associated SNPs did not reveal any significant effect on the SNP

genotypes. Therefore we argue that genetic changes in tumor samples of our cohort may not skew SNP genotype data.

Additionally, the present retrospective study primarily aimed to address two important questions with regard to these risk alleles: firstly, do these alleles also affect the prognosis of the patients and, secondly, do these alleles affect the expression of adjacent genes?

To address the first question, whether these SNPs were associated with the natural course of the disease, i.e., tumor aggressiveness, we focused in this retrospective study on the length of MFS in patients with lymph node-negative disease who did not receive any systemic adjuvant treatment. As adjuvant systemic therapy has a confounding effect on the development of metastases, a population like this one is the most optimal to address such question. Of all the risk alleles investigated, only homozygosity for risk alleles of rs2107425 (AA) in the imprinted gene *H19* was significantly associated with poor prognosis in both univariate and multivariate analysis. To our knowledge, this is the first report showing the pure prognostic value of one of these low-risk loci[1]. However, the underlying biological mechanism accounting for this association is not yet evident. Since this SNP near *H19* does not affect *H19* mRNA expression, it is unlikely that this risk allele and prognostic SNP operates through altering *H19* mRNA expression. This locus is located within the well-known imprinted *IGF2*/*H19* locus on chromosome 11, which contains various imprinted transcripts[22-23]. Loss of imprinting and/or altered expression of other transcripts in this locus may be causal. Lack of an association of the other loci with prognosis could be related to lack of power of our study, but may also indicate that most risk alleles only relate to onset of breast cancer and not to its progression once established. One limitation of our study is that SNP genotype analysis was performed on tumor and not on germline DNA, which is not available for this cohort, and that somatic acquired genetic changes in tumor tissues might affect SNP genotype analysis. To address this, we analyzed a subset of the cohort from which we also had copy number variation data available[21]. In this subgroup, we found no evidence that copy number variation in the regions containing risk-associated SNPs significantly affected SNP genotype frequency. Therefore, we have no evidence that genetic changes in tumor samples of our cohort have skewed the presented SNP genotype data

With regard to the second hypothesis that whether these loci modulate the expression of the nearest genes, we measured transcript expression levels of genes located nearby the risk loci in 1,401 primary breast tumors. For transcript expression analysis, we only focused on the most plausible, non-hypothetical and the nearest genes to the haplotype blocks containing risk-associated loci. The rs2981582 SNP located in the intronic region of the *FGFR2* gene was not related to differences in *FGFR2* mRNA expression. This is puzzling as the *FGFR2* gene has been shown to be amplified and over-expressed in 5-10% of breast cancer patients[24] and because the risk allele itself alters binding affinity for transcription factors of Oct-1/Runx2 and C/EBPβ, which has been causally linked with increased expression of *FGFR2* in breast cancer cell lines[7]. However, such findings were not seen by Sun *et al*, who reported an association of the normal allele with higher *FGFR2* expression in 81 samples[8]. Of note, Sun *et al* analyzed normal breast tissue and not tumor tissue[8].

Altogether, it is unlikely that this SNP acts through altering *FGFR2* mRNA level in breast tumor tissue.

We did not reveal an association of *MYC* transcript expression or *MYC* amplification with SNP genotype rs13281615 located in 8q24. Rs13281615 has no annotated gene in its vicinity. Nevertheless, a recent study showed a tissue-specific physical interaction of enhancers in the 8q24 region with *MYC* expression[16]. However, our analysis does not provide further support for a link between *MYC* mRNA expression and this risk allele. We also measured transcript levels of the class 5 homeobox transcription factor *POU5F1P1*, the nearest gene to this locus, which was recently shown to encode a functional nuclear transcriptional activator[25] and to be over-expressed in prostatic carcinoma[26]. With our assay we could not establish a significant correlation between *POU5F1P1* mRNA expression and the rs13281615 genotype in 1,401 primary breast tumors. It should however be noted that our assay as recently indicated by Applied Biosystems also detects the pseudogene *POU5F1P3* located on chromosome 12, and *POU5F1*, an intronless gene located on chromosome 6 that can encode a protein highly similar to POU class 5 homeobox 1.

For all other loci, we also did not observe any correlation of mRNA expression of the nearest genes with the genotype of the corresponding SNPs (Figure 5.2), except for the risk loci (rs12443621 and rs3803662) on 16q12**.** These SNPs reside in the LD block containing the 5′ end of the *TOX3* gene and a hypothetical gene, *LOC643714*. A recent fine mapping of the LD block has further narrowed this risk-associated locus down to a region of 133 kb spanning the entire coding part of *LOC643714* and a small intergenic region. Moreover, this region exhibits evolutionary conservation and open chromatin conformation suggesting a regulatory function[27]. Our study shows that the alleles in this locus with strongest association with risk (rs3803662 and rs12443621) were significantly related with a lower expression of *TOX3* mRNA. The lower expression of the *TOX3* gene, which encodes a high mobility group box nuclear protein involved in calcium dependent transcription and structural modification of chromatin, in tumors with risk alleles of rs3803662, suggests that *TOX3* might act as a tumor suppressor gene. It is important to note that although the risk alleles of rs3803662 are significantly associated with lower expression of *TOX3,* yet another recent study showed an association of the rs3803662 SNP with an increased expression of the *RBL2* gene, which is also located nearby this risk allele[27]. Although this study including only 77 breast tumors is underpowered compared with our analysis, it might however be possible that this locus differentially regulates more than one distant gene in *cis* or in *trans*.

In summary, this study for the first time estimates the pure prognostic value of SNPs that are associated with low breast cancer risk. We have found that homozygous risk alleles of rs2107425 SNP were associated with an aggressive course of the disease. Secondly, we associated these low risk SNPs with expression of nearby genes. Based on our findings that the risk alleles of rs3803662 and rs12443621 near the *TOX3* gene were associated with a lower expression of *TOX3* mRNA, we hypothesize a tumor suppressor role of this gene. However, functional studies to determine reliably the role of the *TOX3* gene in both normal and malignant settings are required. Finally, the breast cancer risk posed by rs2981582 SNP in *FGFR2* and rs13181615 in the 8q24 region may not operate through the expression modulation of *FGFR2* and *MYC* genes, respectively.

ACKNOWLEDGEMENTS

We are especially grateful to Pascal Arp, Iris van Staveren, Marion Meijer-van Gelder and Mieke Timmermans for their contribution and technical support. The authors thank also the surgeons, pathologists, and internists of the St. Clara Hospital, Ikazia Hospital, St. Franciscus Gasthuis at Rotterdam, and Ruwaard van Putten Hospital at Spijkenisse for the supply of tumor tissues and/or their assistance in the collection of clinical follow-up data. This study was in part financed by the Netherlands Genomics Initiative (NGI) / Netherlands Organization for Scientific Research (NWO). MR was funded by the Higher Education Commission (HEC) government of Pakistan.

CONFLICT OF INTEREST STATEMENT

All authors declared that they have no competing interests.

SUPPLEMENTARY DATA

Supplementary Tables S5.1, S5.2, S5.3 are available on:
https://www.dropbox.com/sh/4fwrk0lit91dsp5/WP1sZYZAQw?m

REFERENCES

1. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
2. Stacey, S.N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-9 (2007).
3. Turnbull, C. et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-7 (2010).
4. McInerney, N. et al. Low penetrance breast cancer predisposition SNPs are site specific. *Breast Cancer Res Treat* **117**, 151-9 (2009).
5. Hemminki, K. et al. Low-risk variants FGFR2, TNRC9 and LSP1 in German familial breast cancer patients. *Int J Cancer* **126**, 2858-62 (2010).
6. Liang, J. et al. Genetic variants in trinucleotide repeat-containing 9 (TNRC9) are associated with risk of estrogen receptor positive breast cancer in a Chinese population. *Breast Cancer Res Treat* **124**, 237-41 (2010).
7. Meyer, K.B. et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* **6**, e108 (2008).
8. Sun, C., Olopade, O.I. & Di Rienzo, A. rs2981582 is associated with FGFR2 expression in normal breast. *Cancer Genet Cytogenet* **197**, 193-4 (2010).
9. McShane, L.M. et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Res Treat* **100**, 229-35 (2006).

10. Foekens, J.A. et al. Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols. *Cancer Res* **49**, 5823-8 (1989).
11. Berns, E.M. et al. Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumours: correlation with steroid receptors. *Eur J Cancer* **28**, 697-700 (1992).
12. Hsieh, S.M., Look, M.P., Sieuwerts, A.M., Foekens, J.A. & Hunter, K.W. Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. *Breast Cancer Res* **11**, R75 (2009).
13. Sieuwerts, A.M. et al. How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin Cancer Res* **11**, 7311-21 (2005).
14. van Agthoven, T. et al. Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J Clin Oncol* **27**, 542-9 (2009).
15. Mavaddat, N., Antoniou, A.C., Easton, D.F. & Garcia-Closas, M. Genetic susceptibility to breast cancer. *Mol Oncol* **4**, 174-91 (2010).
16. Ahmadiyeh, N. et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A* **107**, 9742-6 (2010).
17. Gold, B. et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A* **105**, 4340-5 (2008).
18. Thomas, G. et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-84 (2009).
19. Garcia-Closas, M. et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**, e1000054 (2008).
20. Milne, R.L. et al. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J Natl Cancer Inst* **101**, 1012-8 (2009).
21. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-8 (2009).
22. Tilghman, S.M. The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell* **96**, 185-93 (1999).
23. Tycko, B. & Efstratiadis, A. Genomic imprinting: piece of cake. *Nature* **417**, 913-4 (2002).
24. Adnane, J. et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* **6**, 659-63 (1991).
25. Panagopoulos, I., Moller, E., Collin, A. & Mertens, F. The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. *Oncol Rep* **20**, 1029-33 (2008).
26. Kastler, S. et al. POU5F1P1, a putative cancer susceptibility gene, is overexpressed in prostatic carcinoma. *Prostate* **70**, 666-74 (2010).
27. Udler, M.S. et al. Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet* **19**, 2507-15 (2010).

Chapter 6

High *TWIST1* mRNA expression is associated with poor prognosis in lymph node-negative and estrogen receptor-positive human breast cancer and is co-expressed with stromal as well as ECM related genes

Muhammad Riaz, Anieta M Sieuwerts, Maxime P Look, Mieke A Timmermans, Marcel Smid, John A Foekens, John WM Martens.

Breast Cancer Res. 2012; 14(5):R123

ABSTRACT

Introduction TWIST homolog 1 (*TWIST1*) is a transcription factor that induces epithelial to mesenchymal transition (EMT), a key process in metastasis. The purpose of this study was to investigate whether *TWIST1* expression predicts disease progression in a large breast cancer cohort with long-term clinical follow-up, and to reveal the biology related to *TWIST1* mediated disease progression.

Patients and Methods *TWIST1* mRNA expression level was analyzed by quantitative real-time reverse polymerase chain reaction (RT-PCR) in 1,427 primary breast cancers. In uni- and multivariate analysis using Cox regression, *TWIST1* mRNA expression level was associated with metastasis-free survival (MFS), disease-free survival (DFS) and overall survival (OS). Separate analyses in lymph node-negative patients (LNN, n = 778) who did not receive adjuvant systemic therapy, before and after stratification into estrogen receptor (ER)-positive (n = 552) and ER-negative (n = 226) disease, were also performed. The association of *TWIST1* mRNA with survival endpoints was assessed using Kaplan-Meier analysis. Using gene expression arrays, genes showing a significant Spearman rank correlation with *TWIST1* were used to identify overrepresented Gene Ontology (GO) terms and KEGG-annotated biological pathways.

Results Increased mRNA expression level of *TWIST1* analyzed as a continuous variable in both uni- and multivariate analysis was associated with shorter MFS in all patients (hazard ratio (HR): 1.17, 95% confidence interval, (95% CI):1.09-1.26; and HR: 1.17, 95% CI:1.08-1.26; respectively), in LNN patients (HR: 1.22, 95% CI:1.09-1.36; and HR: 1.21, 95% CI:1.07-1.36; respectively) and in the ER-positive subgroup of LNN patients (HR: 1.34, 95% CI:1.17-1.53; and HR: 1.32, 95% CI:1.14-1.53; respectively). Similarly, high *TWIST1* expression was associated with shorter DFS and OS in all patients and in the LNN/ER-positive subgroup. In contrast, no association of *TWIST1* mRNA expression with MFS, DFS or OS was observed in ER-negative patients. Genes highly correlated with *TWIST1* were significantly enriched for cell adhesion and ECM-related signaling pathways. Furthermore, *TWIST1* mRNA expression was highly in tumor stroma and positively related to tumor stromal content ($P < 0.001$).

Conclusions *TWIST1* mRNA expression is an independent prognostic factor for poor prognosis in LNN/ER-positive breast cancer. The biological associations suggest an involvement of the tumor microenvironment in *TWIST1*'s adverse role in breast cancer.

INTRODUCTION

Breast cancer is one of the most frequently diagnosed cancers and the leading cause of cancer related deaths among females of the Western world[1]. Patients do not die from the primary tumor, but from metastases, which already are resistant or acquire resistance to systemic therapy. Metastasis is a complex multi-step process in which malignant cells undergo sequential molecular changes helping them to disengage from primary sites, intravasate into blood vessels, extravasate to distant organs and finally colonize secondary sites. Each of these metastatic steps is affected by aberrant expression of a variety of transcription factors and among them, TWIST homologue 1 *(TWIST1)* is considered an important regulator of disease progression[2].

TWIST1 protein, encoded by the *TWIST1* gene, is a member of a large protein family called basic helix-loop-helix (bHLH) transcription factors[3]. Most family members contain a bHLH domain, which enables it to target specific DNA sequences and thereby allowing them to regulate developmental processes in many organs and tissues. *TWIST1* plays a key role in the regulation of embryogenesis, gastrulation, and mesoderm formation during early embryonic development of Drosophila and many other species[4,5]. An autosomal mutation pattern in the *TWIST1* gene leads to Saethre-Chotzen syndrome, a genetic condition characterized by premature fusion of skull bones affecting symmetrical growth of head and face[6]. In children, TWIST1 protein is involved in adequate maturation of the skull and spine bones and normal development of arms and hind legs.

More recently, TWIST1 protein has been implicated in various carcinomas including breast cancer, where it plays a role in metastasis through activation of a biologically latent developmental process called epithelial to mesenchymal transition (EMT)[7,8]. In the EMT process, malignant epithelial cells undergo cytoskeletal changes including the down-regulation of epithelial markers such as E-cadherin and co-expressed catenins and up-regulation of mesenchymal markers such as vimentin, N-cadherin, and fibronectin. EMT transformed malignant cells are more motile and can be more efficient in invading the surrounding tissues and as a result metastasize to distant organs[9].

In this large retrospective study of 1,427 primary breast cancer patients, we determined whether *TWIST1* gene expression level is a prognostic marker. To avoid possible confounding effects of therapy and to study the natural course of the disease, we particularly focused on the subgroup of 778 lymph node-negative (LNN) patients who did not receive any adjuvant systemic therapy. Additionally, to understand the biological context of TWIST1, we have identified, using available Affymetrix U133A gene expression data[10,11] genes and biological pathways co-expressed with *TWIST1*. By doing this, we identified a clear link between the tumor microenvironment and *TWIST1* expression in clinical breast cancer.

PATIENTS AND METHODS

Patients

The study was approved by the institutional medical ethics committee of the Erasmus MC (MEC 02.953). The study was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands[12], informed consent was not required. Wherever possible, the study has been reported in line with the Reporting Recommendations for Tumor Marker Prognostic Studies guidelines[13,14]. Breast cancer tissue specimens from 1,427 female patients with primary operable breast cancer were included in the study. Tumor samples were originally submitted to our reference laboratory from 25 regional hospitals for biochemical assessment of steroid hormone receptor status. Guidelines for primary treatment were similar in all hospitals. All available frozen breast tumor samples of at least 100 mg from female patients with invasive breast cancer who entered the clinic during the period of 1978 to 2001, and from whom detailed clinical follow-up data was available were processed.

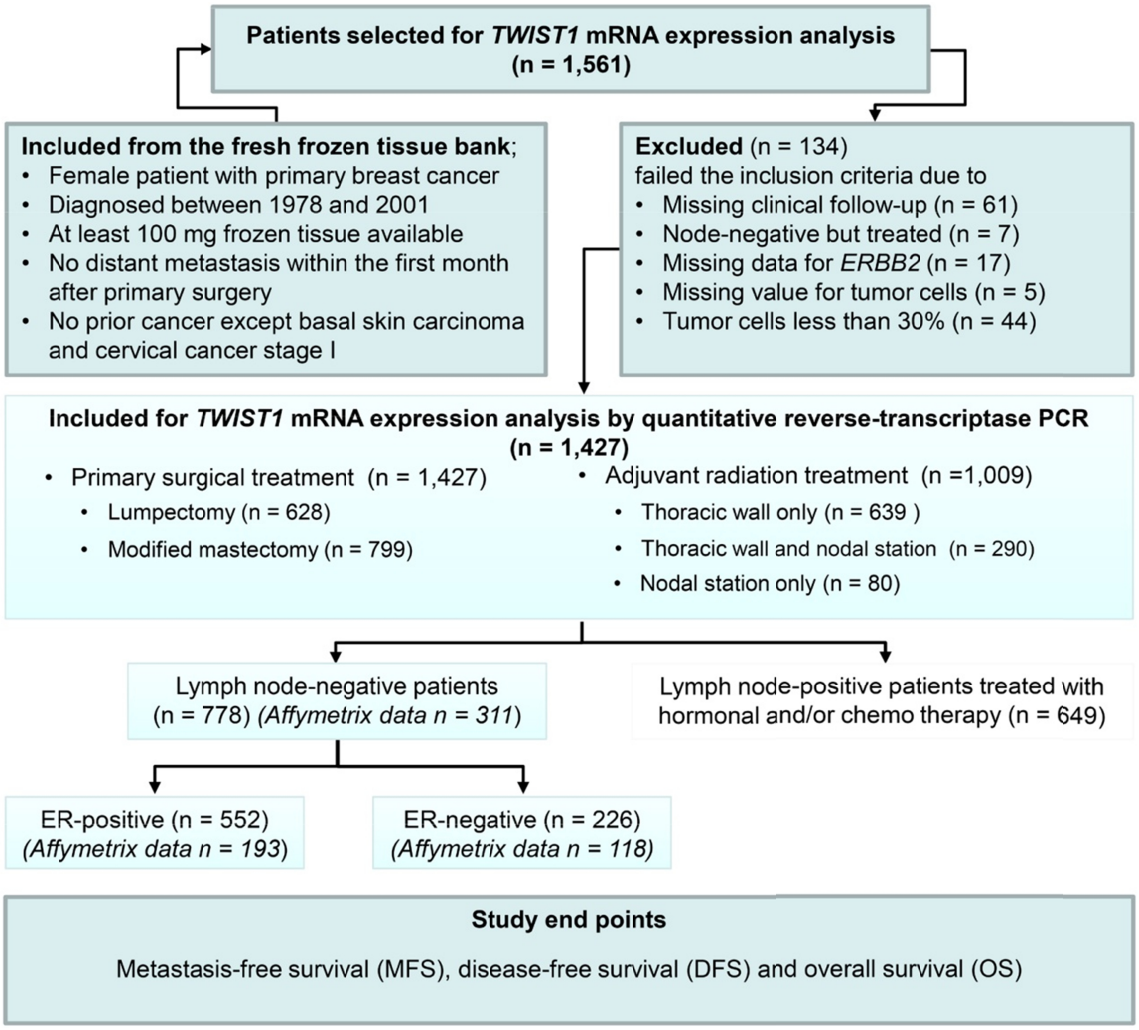


Figure 6.1 Study design, patient inclusion criteria, and patient sub cohorts analyzed. Associations of *TWIST1* mRNA expression levels with tumor aggressiveness were evaluated in all patients and in LNN, LNN/ER-positive, and LNN/ER-negative patients. The LNN patients in this study did not receive adjuvant systemic therapy.

Moreover, these patients had no distant metastasis within first month after primary surgery and no prior cancer except basal skin carcinoma and cervical cancer stage I. ER, PGR and HER2 (ERBB2) levels were assessed by quantitative real-time reverse transcriptase PCR (qRT-PCR) as described before[15-17]. Lymph node involvement, tumor size and grade were extracted from the pathology reports as obtained from the hospitals. Primary surgical treatment was lumpectomy in 628 patients (44%) or modified mastectomy for 799 patients (56%). One thousand and nine patients (71%) received adjuvant radiotherapy at the thoracic wall only (n = 639) and either at the thoracic wall and nodal stations (n = 290) or at the nodal stations only (n = 80). Thirty-three percent of the patients had T1 tumors. The median age of the patients at surgery was 55 years (range, 23-89 years). Routine post-surgical follow-up and the definition of the endpoints of MFS, DFS and OS were described before[11,18]. During follow-up, a local/regional relapse for 127 patients was not counted as an event in the analysis of MFS. The median follow-up time was 104 months (range, 4-262 months) with 707 and 669 events in the analyses of MFS and OS, respectively. Study design, patient inclusion criteria, and patient subgroups included in the analyses are given in Figure 6.1. Other relevant clinico-pathological characteristics are listed in Table 6.1.

Table 6.1 Association of *TWIST1* mRNA expression with standard clinical and pathological characteristics of patients and tumors

Characteristics	**No. patients**	***TWIST1* mRNA expression** (Median value*)	**P-value**
Age (years)			
≤ 40	188	0.025	
41-55	530	0.028	
56-70	471	0.030	
> 70	238	0.020	0.04†
Menopausal status			
Premenopaual	603	0.027	
Postmenopausal	824	0.027	0.21‡
Tumor size			
pT1, ≤ 2 cm	463	0.031	
pT2, > 2-≤ 5	807	0.025	
pT3, > 5 + pT4	157	0.028	0.009¶
Lymph nodes involved			
0	778	0.026	
1-3	288	0.025	
>3	361	0.030	0.007¶
Grade			
Poor	800	0.026	
Good/Moderate	230	0.027	
Unknown	397	0.029	0.47¶
Tumor histology			
IDC	957	0.027	
ILC	112	0.044	<0.001‡, §
Medullary	31	0.016	
Mucinous	39	0.008	<0.001¶
ER status (mRNA)			
Negative	360	0.021	
Positive	1,067	0.029	0.06†

Table 6.1 Continued

Characteristics	No. patients	*TWIST1* mRNA expression (Median value*)	P-value
PGR status (mRNA)			
Negative	572	0.023	
Positive	855	0.029	<0.001†
ERBB2 status (mRNA)			
Negative	1,187	0.026	
Positive	240	0.032	<0.001†

Abbreviation: IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma.
* Expression values after normalization to our set of 3 reference genes (*B2M*, *HMBS* and *HPRT1*).
† P Spearman rank correlation test.
‡ P for Mann-Whitney test.
¶ P for Kruskal-Wallis test.
§ *TWIST1* expression was compared between IDC and ILC tumors.

Tissue processing and estimation of the amount of epithelial tumor cells

Tissue processing was done as described in detail before[15,19]. In brief, 20 to 60 cryostat sections of 30 µm size corresponding to 30 to 100 mg weight were cut from frozen tissues. Three 5 µm sections were cut before, in between, and after cutting the sections for RNA isolation, and stained with hematoxylin and eosin (H&E) to assess the amount of tumor cells relative to the amount of surrounding stromal cells. The amount of nuclei evidently of epithelial tumor cell origin relative to the amount of surrounding stromal cells was estimated with a 100-fold magnification in 10 different areas covering the area of each of the three H&E sections. Only specimen with at least 30% of the nuclei of epithelial tumor cell origin and distributed uniformly over at least 70% of the section area were included. Like before[16], these estimates were used to dichotomize our tumor cohort at the median level of 70% tumor cell nuclei in stromal-rich (SR) (primary tumors containing $\geq$30% stromal components) and stromal-poor (SP) (primary tumors containing at least 70% tumor cells).

Laser capture microdissection

Laser capture microdissection (LCM) was performed according to a previously published method[20]. Briefly, 8 µm sections of a fresh frozen tumor tissue were cryosectioned and pasted on UV sterilized polyethylene naphthalate (PEN) covered glass slides (P.A.L.M. Microlaser Technologies, Bernried, Germany). The slides were air dried for 10 sec at room temperature, fixed in ice-cold 70% ethanol, and washed shortly in Milli-Q water. The slides were then stained for 15 sec in Haematoxylin (Klinipath 4085.9005, Duiven, The Netherlands) and washed again in Milli-Q water. Subsequently, the slides were dehydrated twice in RNase-free 50, 70, 95, and 100% ethanol for 30 sec each, and air-dried. Next, LCM was performed directly on the stained sections. From consecutive cryosections of a tumor tissue, both the stromal content and tumor epithelial cells were collected separately in triplicate in P.A.L.M. tube caps containing 25µl of tissue lysis (RLT+) buffer (Qiagen 1053393, Venlo, The Netherlands) using a P.A.L.M. LCM device, type P-MB (P.A.L.M. Microlaser Technologies AG, Bernried, Germany). Both stromal and epithelial cell fractions were then spun down into 0.5-ml

Eppendorf Protein LoBind tubes (Eppendorf, Hamburg, Germany) containing 225µl RLT+ buffer and used for RNA isolation.

RNA isolation, cDNA synthesis, and quantification of specific mRNA species

RNA isolation, cDNA synthesis, quantification of specific mRNA species, and quality control checks were done as described previously by Sieuwerts et al[15]. qRT-PCR was performed in 25-µL reaction volume in a Mx3000P Real-time PCR System (Agilent, Amsterdam, the Netherlands) using a commercially available gene expression assays for *TWIST1* (Hs00361186_m1) and *VIM* (Hs Hs00185584_m1) from Applied Biosystems/Life Technologies (Nieuwerkerk aan den IJssel, The Netherlands). For *EPCAM* mRNA measurement, a combination of the following primers and probe was used (F: 3'-AGT TTG CGG ACT GCA CTT CA, R: 3'- AAT ACT CGT GAT AAA TTT TGG ATC CA, FAM-labeled MGB probe: AAG GAG ATC ACA ACG CGT). Primer sequences for estrogen receptor-α (*ESR1*), progesterone receptor (*PGR*), *ERBB2* and the reference genes (*HMBS*, *HPRT1* and *B2M*), as well as how PCR reactions and validations were carried out to ensure PCR specificity, were done as described before[15,17]. Levels of mRNA expression were quantified relative to the set of reference genes as described before[15].

Immunohistochemistry (IHC)

Sections of 4 µm were cut from formalin-fixed paraffin-embedded tissues, pasted on glass slides, and dried overnight at 37°C. The slides were dewaxed in xylene solution and hydrated in series of different percentages of alcohol (100%, 95%, 75%, and 50%). The antigen was retrieved for 40 min in a hot water bath (at 95°C) in Tris-EDTA buffer, pH 9.0 (DAKO; S2367, Glostrup, Denmark) and cooled down for 20 min at room temperature. After retrieval, slides were immersed in 3% hydrogen peroxide for 10 min to block endogenous peroxidase activity, and in a DAKO protein-free serum blocking solution for 30 minutes to block non-specific binding sites. The slides were then incubated overnight at 4°C with a mouse monoclonal primary antibody raised against a human recombinant fragment of TWIST1 (AbCam TWIST2C1a, ab50887, Cambridge, UK), diluted 1:100 in an antibody diluent (DAKO). Negative controls were made by replacing the primary antibody with mouse immunoglobulin at an appropriate dilution. Next, the slides were treated with a secondary antibody (envision mouse kit, DAKO) and visualized using DAB (envision kit, DAKO). The slides were counterstained with hematoxylin and dehydrated through graded alcohol and xylene. Evaluation of the immunostaining and histoscoring were performed according to a previously published method [21]. Briefly, a histoscore of 0-300 was given to each tumor by multiplying the proportion of cells with nuclear staining in a tissue by the intensity of the staining as follows; histoscore = (0 × percentage not stained) + (1 × percentage weakly stained) + (2 × percentage moderately stained) + (3 × percentage strongly stained). The histoscores calculated by this way were correlated with *TWIST1* mRNA expression. The intensity of TWIST1 stromal expression in tumor tissues was scored as no, weak, moderate or strong staining.

Identification of *TWIST1* co-expressed genes on Affymetrix U133A gene-chips and identification of Gene Ontology GO) terms and biological pathways

In this analysis we used *TWIST1* mRNA expression, which was previously measured by Affymetrix U133A gene-chips (Affymetrix, Santa Clara, CA, USA) (probe set 213943_at)[10,11,22]. The mRNA expression levels of *TWIST1* measured on the Affymetrix-array and by qRT-PCR were highly correlated in an overlap of 193 LNN/ER-positive and 118 LNN/ER-negative tumor specimens using the Spearman-Rank correlation test (Spearman rank correlation, Rs = 0.7; *P* <0.001 for both). Genes showing a significant positive or negative correlation with *TWIST1* were identified by using these Affymetrix-array data after all genes were corrected for multiple testing by applying a false discovery rate (FDR) of 5%[23]. Biological pathways and GO terms from genes significantly co-expressed with *TWIST1* were identified using the ArrayTrack™ software (NCTR/FDA, Jefferson, AR 72079, USA)[24]. The overrepresented pathways and GO terms identified by ArrayTrack were based on Fisher's exact statistics. Details about how ArrayTrack™ software functions can be retrieved from the website[25]. The gene expression data used have previously been deposited in the National Center for Biotechnology Information/Gene Expression Omnibus database entries GSE2034 and GSE5327.

Statistics

Computations were done with the use of the STATA statistical package, release 11.2 (STATA Corp. College Station. TX). Differences in levels were assessed with the Mann-Whitney U test or Kruskal-Wallis test. In these tests, patient and tumor characteristics were used as grouping variables. The strengths of the associations between continuous variables were tested with the Spearman rank correlation test. *TWIST1* mRNA expression levels measured by qRT-PCR were log transformed to reduce the skewness (Skewness-Kurtosis normality test, *P* <0.05) and to attain symmetric distribution (Additional file 6.1, Figure S1). The prognostic value of the clinical and biological variables, with MFS, OS, and DFS as the end points in the univariate and multivariate analyses, were investigated with the use of the Cox proportional hazard model. The hazard ratio (HR) and its 95% confidence interval (95% CI) were derived from these models by analyzing *TWIST1* mRNA expression levels as a continuous variable or after dividing their levels into quartiles. The proportionality assumption was investigated with a test based on the Schoenfeld residuals. Kaplan-Meier survival analysis was performed and a log-rank test was used to assess a trend of survivor's function across four quartiles of *TWIST1* mRNA expression levels. All P-values are two sided, and *P* <0.05 was considered statistically significant.

RESULTS

Association of *TWIST1* mRNA expression with standard clinical and pathological characteristics of patients and tumors

Association of *TWIST1* mRNA expression levels with clinical and pathological characteristics of patients and tumors are shown in Table 6.1. Tumor *TWIST1* mRNA expression was higher in patients

with >3 positive lymph nodes, and lower in pT2 tumors and in older patients. Furthermore, *TWIST1* mRNA expression was higher in invasive lobular carcinoma (ILC) compared with invasive ductal carcinoma (IDC). Mucinous and medullary-type of tumors showed the lowest *TWIST1* mRNA expression levels. In addition, *TWIST1* mRNA expression level was higher in *PGR*-positive and *ERBB2*-overexpressing tumors, but was not associated with tumor grade and menopausal status.

Association of *TWIST1* mRNA expression with prognosis in primary breast cancer patients

To explore the prognostic significance of *TWIST1* mRNA expression in primary breast cancer patients, we first performed Cox uni- and multivariate analyses (traditional factors in the multivariate analysis were; age, menopausal status, number of positive lymph nodes, tumor size, tumor grade, ER status, PGR status and ERBB2 status) for MFS, DFS and OS as a function of continuous *TWIST1* mRNA expression levels. In all 1,427 patients, increasing *TWIST1* mRNA expression levels were associated with shorter MFS in both uni- and multivariate analysis (HR: 1.17, 95% CI: 1.09-1.26; *P* <0.001 and HR: 1.17, 95% CI: 1.08-1.26; *P* <0.001; respectively) (Table 6.2).

Table 6.2 Model for Uni-and Multivariate analysis for metastasis-free survival in all patients (n = 1,427)

Factors	No. patients	Univariate			Multivariate		
		HR	95% CI	P-values	HR	95% CI	P-values
Age (years)							
≤ 40	188	1			1		
41-55	530	0.92	0.73-1.16	0.47	0.88	0.69-1.12	0.30
56-70	471	1.02	0.80-1.29	0.88	0.85	0.59-1.21	0.36
> 70	238	0.80	0.66-1.06	0.12	0.69	0.47-1.03	0.07
Menopausal status							
Premenopaual	603	1			1		
Postmenopausal	824	1.07	0.92-1.25	0.35	1.23	0.94-1.62	0.14
Tumor size							
pT1, ≤ 2 cm	463	1			1		
pT2, > 2-≤ 5	807	1.65	1.39-1.97	<0.001	1.38	1.15-1.65	<0.001
pT3, > 5 + pT4	157	2.80	2.20-3.58	<0.001	1.80	1.39-2.33	<0.001
Lymph nodes involved							
0	778	1			1		
1-3	288	1.79	1.48-2.17	<0.001	1.71	1.41-2.08	<0.001
>3	361	2.92	2.46-3.46	<0.001	2.63	2.19-3.16	<0.001
Grade							
Poor	800	1			1		
Good/Moderate	230	0.51	0.40-0.65	<0.001	0.64	0.50-0.82	<0.001
Unknown	397	0.89	0.75-1.05	0.17	0.99	0.83-1.17	0.87
ER status (mRNA)							
Negative	360	1			1		
Positive	1,067	0.82	0.70-0.97	0.02	0.83	0.67-1.04	<0.001
PGR status (mRNA)							
Negative	572	1			1		
Positive	855	0.75	0.64-0.87	<0.001	0.87	0.71-1.06	0.17
ERBB2 status (mRNA)							
Negative	1,187	1			1		
Positive	240	1.30	1.08-1.57	0.006	1.18	0.97-1.43	0.09

Table 6.2 Continued

Factors	No. patients	Univariate			Multivariate		
		HR	95% CI	P-values	HR	95% CI	P-values
					Addition to the base model		
***TWIST1* mRNA level†** (Continuous)	1,427	1.17	1.09-1.26	<0.001	1.17	1.08-1.26	<0.001
***TWIST1* mRNA level†**							
TWIST1 quartile 1	360	1			1		
TWIST1 quartile 2	355	1.16	0.94-1.45	0.17	1.09	0.88-1.37	0.43
TWIST1 quartile 3	357	1.22	0.98-1.51	0.08	1.25	1.00-1.56	0.05
TWIST1 quartile 4	355	1.49	1.20-1.83	<0.001	1.43	1.15-1.77	0.001

†*TWIST1* mRNA expression level (log-transformed after normalization to the expression of three reference genes) was separately added to the base multivariate model that included the factors age, menopausal status, tumor size, grade, lymph-node status, *ER*, *PGR*, and *ERBB2* mRNA expression levels, tested both as a continuous variable and in four quartiles.

To visualize the prognostic value of *TWIST1* mRNA expression in Kaplan-Meier survival curves, we divided *TWIST1* expression levels into four quartiles Q1 (low) to Q4 (high) (Figure 6.2). To exclude a possible confounding treatment effect, we also evaluated the prognostic value of *TWIST1* mRNA expression in LNN-patients who did not receive any adjuvant systemic therapy. We also stratified these LNN patients into ER-positive and ER-negative subgroups as these subgroups are considered biologically distinct. In the analysis of all 778 LNN patients, an association of increasing *TWIST1* mRNA expression levels with shorter MFS was observed in both uni- and multivariate analysis (univariate HR: 1.22, 95% CI: 1.09-1.36; *P* =0.001 and HR: 1.21, 95% CI: 1.07-1.36; *P* =0.001; respectively) (Additional file 6.2, Table S1). Moreover, within the LNN subgroup of patients, the association of increasing *TWIST1* mRNA levels with shorter MFS was confined to the ER-positive subgroup of patients [univariate (HR: 1.34, 95% CI: 1.17-1.53; *P* <0.001) and multivariate (HR: 1.32, 95% CI: 1.14-1.53; *P* <0.001)] (Table 6.3). There was no association of *TWIST1* mRNA levels with MFS in the ER-negative subgroup of patients (univariate HR = 1.05, 95% CI: 0.86-1.29; *P* =0.64) (Additional file 6.2, Table S2).

Kaplan-Meier survival curves for all patients and subgroups of patients with LNN and LNN/ER-positive and LNN/ER-negative disease, after dividing these cohorts in quartiles (Q1 to Q4), are shown in Figure 6.2A-D. Cox regression analysis for MFS showed that in the LNN/ER-positive group the quartile with the highest *TWIST1* mRNA levels (Q4) had a 2-fold increased hazard ratio compared with the quartile with the lowest *TWIST1* mRNA levels (Q1) (Table 6.3).

We also related *TWIST1* mRNA expression levels with OS in all patients, in the LNN cases as a whole and after stratifying by ER status. Similarly as observed for MFS, increasing *TWIST1* mRNA expression levels were associated with shorter OS in all patients (univariate HR: 1.10, 95% CI: 1.08-1.26, *P* =0.013; multivariate HR: 1.11, 95% CI: 1.03-1.20, *P* =0.006), in the LNN (univariate HR: 1.13, 95% CI: 1.01-1.27, *P* = 0.03; multivariate HR: 1.16, 95% CI: 1.03-1.30, *P* =0.01) and in the LNN/ER-positive group of patients (univariate HR: 1.26, 95% CI: 1.09-1.45, *P* =0.001; multivariate HR: 1.27, 95% CI: 1.11-1.46, *P* =0.001). These associations were not significant in the LNN/ER-negative group of patients (univariate HR: 0.96, 95% CI: 0.78-1.18, *P* =0.69; multivariate HR: 0.98, 95% CI: 0.79-1.21, *P* =0.85). The results are included in Additional file 6.2, Tables S3-6. Kaplan-Meier survival curves

showing OS for all patients, and also for the subgroups of patients after dividing them into quartiles based on *TWIST1* mRNA levels are shown in Additional file 6.1, Figure S2A-D.

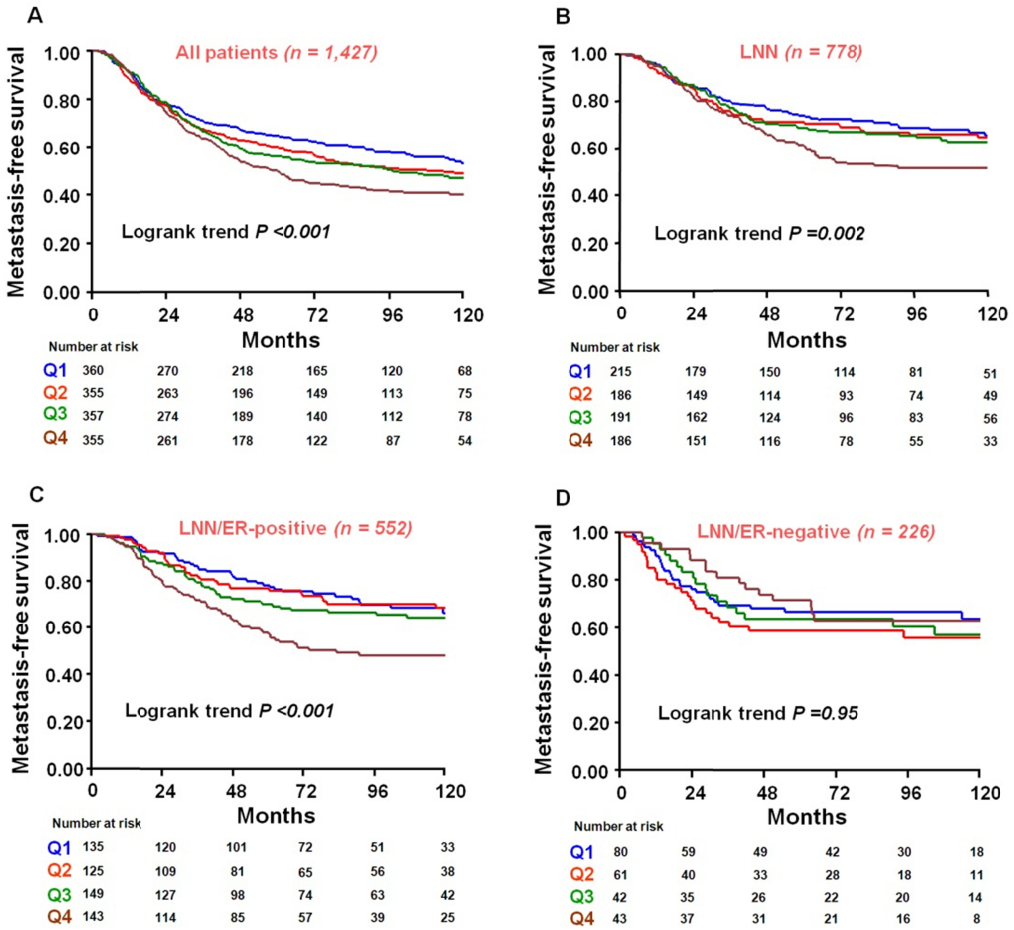


Figure 6.2 Kaplan-Meier survival curves presenting the association of *TWIST1* mRNA expression with metastasis-free survival in; **A:** all 1,427 patients; **B:** LNN patients only, **C:** LNN and ER-positive patients and **D:** LNN and ER-negative patients. The LNN patients in this study did not receive adjuvant systemic therapy. The patients are divided into four quartiles [Q1 (low) to Q4 (high)] based on *TWIST1* mRNA expression levels after normalizing these to the expression of our set of 3 reference genes (*B2M, HMBS* and *HPRT1*). Patients at risk at various time points are indicated.

We finally analyzed whether *TWIST1* mRNA expression levels were associated with DFS in all patients and in LNN/ER-positive patients only, and obtained similar results as were obtained for MFS and OS (Additional file 6.2, Tables S7-8).

Table 6.3 Model of Uni- and Multivariate analysis for metastasis-free survival in lymph-node negative and ER-positive patients (n = 552)

Factors	No. patients	Univariate			Multivariate		
		HR	95% CI	P-values	HR	95% CI	P-values
Age (years)							
≤ 40	66	1			1		
41-55	187	0.97	0.64-1.49	0.90	1.05	0.68-1.64	0.81
56-70	179	0.79	0.51-1.23	0.29	0.81	0.40-1.66	0.57
> 70	120	0.59	0.35-0.99	0.05	0.59	0.27-1.28	0.18

Table 6.3 Continued

Factors	No. patients	Univariate			Multivariate		
		HR	95% CI	P-values	HR	95% CI	P-values
Menopausal status							
Premenopaual	215	1			1		
Postmenopausal	337	0.76	0.57-1.00	0.05	0.99	0.56-1.74	0.98
Tumor size							
pT1, ≤ 2 cm	251	1			1		
pT2, > 2-≤ 5	276	1.12	0.84-1.49	0.43	1.09	0.82-1.46	0.34
pT3, > 5 + pT4	25	1.62	0.84-3.11	0.15	1.92	0.99-3.75	0.06
Grade							
Poor	257	1			1		
Good/Moderate	124	0.53	0.35-0.79	0.002	0.56	0.37-0.85	0.007
Unknown	171	1.06	0.78-1.44	0.70	1.20	0.87-1.64	0.26
PGR status (mRNA)							
Negative	116	1			1		
Positive	436	0.67	0.49-0.92	0.013	0.71	0.51-0.99	0.04
ERBB2 status (mRNA)							
Negative	475	1			1		
Positive	77	1.51	1.06-2.15	0.02	1.39	0.97-2.00	0.08
					Addition to the base model		
***TWIST1* mRNA level**† (Continuous)	552	1.34	1.17-1.53	<0.001	1.32	1.14-1.53	<0.001
***TWIST1* mRNA level**†							
TWIST1 quartile 1	135	1			1		
TWIST1 quartile 2	125	1.02	0.65-1.61	0.92	0.90	0.57-1.43	0.67
TWIST1 quartile 3	149	1.28	0.85-1.94	0.24	1.09	0.71-1.68	0.70
TWIST1 quartile 4	143	2.06	1.39-3.04	<0.001	1.87	1.25-2.80	0.002

† *TWIST1* mRNA expression level (log-transformed after normalization to the expression of three reference genes) was separately added to base multivariate model that included the factors age, menopausal status, tumor size, grade, *ER*, *PGR*, and *ERBB2* mRNA expression levels, tested both as a continuous variable and in four quartiles.

Investigation of *TWIST1* related biological pathways in LNN breast cancer

To explore the underlying biology of *TWIST1*-associated disease progression in breast cancer, we determined *TWIST1* co-expressed genes previously measured on Affymetrix U133A gene-chips[10,11]. Since we found *TWIST1* to be associated with poor prognosis in LNN/ER-positive patients only, we first decided to include only the LNN/ER-positive samples to examine the *TWIST1* co-expressed genes. After using a Spearman correlation test followed by applying a multiple testing correction at a FDR of 5%, we found 1,847 genes to be positively correlated and 1,445 genes to be negatively correlated with *TWIST1* mRNA expression (Additional file 6.3, Table S9A, B). To reveal which biological pathways might be associated with *TWIST1* mRNA expression, we performed pathway analyses using the genes co-expressed with *TWIST1* and the KEGG pathway database as input for the analysis in the ArrayTrack™ software package. The three most significantly overrepresented pathways containing these *TWIST1* co-expressed genes were: Focal adhesion pathway (Fisher's exact $P = 1.0\times10^{-8}$), ECM-receptor interaction pathway (Fisher's exact $P = 9.9\times10^{-7}$), and TGF-beta signaling pathway (Fisher's exact $P = 2.6\times10^{-4}$) (Additional file 3, Table S10). In addition, GO term enrichments among *TWIST1* co-expressed genes identified "cell adhesion", "extracellular structure organization", and "anatomical structure formation involved in morphogenesis" as the top three

significant GO terms (level $\geq$ 3; gene hits > 5; *P* values $\leq$0.001) in the category of "biological processes", and the GO terms, "extracellular matrix" proteinaceous extracellular matrix, and "extracellular matrix part" as the top three significant GO terms in the category "protein cellular components" (Additional file 6.3, Table S11).

To evaluate a possible *TWIST1* mediated biological difference between LNN/ER-positive and LNN/ER-negative patients, we also performed pathway analyses using *TWIST1* co-expressed genes (n = 1,141 positively and n = 498 negatively correlated genes with *TWIST1*, respectively, Additional file 6.3, Table S12A & B) measured on Affymetrix arrays from LNN/ER-negative patients (n = 118). As observed in LNN/ER-positive patients, these analyses identified Focal adhesion and ECM-receptor interaction pathways as the two most significant pathways. But, in contrary to what unlike observed in LNN/ER-positive patients, the TGF-beta signaling pathway was not found to be the third most significant pathway in LNN/ER-negative patients (Additional file 6.3, Table S13). We also performed pathway analysis using *TWIST1* co-expressed genes in all LNN patients irrespective of ER status. Analyses revealed only Focal adhesion and ECM-receptor interaction pathways as the two most significant pathways but TGF-beta signaling pathway was not found to be as the third top significant pathway in all patients (Additional file 6.3, Table S14). Altogether, these data suggest that the co-expression of TGF-beta signaling pathway with *TWIST1* gene is more prominent in LNN/ER-positive cancer.

Association of *TWIST1* mRNA expression levels with stromal content

The co-expression analysis suggested an association between stromal content of the tumor tissue and *TWIST1* mRNA expression levels. Therefore, we studied the association of *TWIST1* mRNA expression with tumor stromal content in more detail. For this, we performed LCM to separately isolate the stromal (S) and epithelial tumor cell (T) areas from 10 breast tumor tissues (n = 9 *ESR1* high, n = 1 *ESR1* low) with sufficiently micro-dissectible fresh frozen tissue left. After LCM we measured *TWIST1* mRNA by qRT-PCR in RNA isolated from both the stromal and the epithelial tumor cell fractions of these tissues, and normalized the expression levels to our set of reference genes. To ensure that our reference gene-set did not cause a bias in these analyses, and our LCM material was indeed enriched for stromal and epithelial cells, we also measured *VIM* and *EpCAM* (Epithelial Cell Adhesion Molecule) in these fractions (Figure 6.3A). In these analyses, *EpCAM* mRNA expression was high in the epithelial tumor cell fractions [3.4-fold, 2-sided paired t-test *P* =0.006] and *VIM* and *TWIST1* mRNA expression were high in the stromal fractions [23.4-fold, *P* =0.0001 and 14.6-fold, *P* =0.0004, respectively].

As the prognostic value for *TWIST1* mRNA expression is confined to ER-positive breast cancer, we assessed whether ER expression might be of relevance for *TWIST1* mRNA expression in tumor tissues with a variable amount of stromal content. For this, we compared *TWIST1* mRNA expression levels between stromal rich (SR) and stromal poor (SP) tumors in the ER-positive and ER-negative subgroups. Notably, only in ER-positive tumors, *TWIST1* mRNA expression was significantly higher in SR compared with SP tumors (Figure 6.3B, *P* <0.001); in ER-negative tumors the difference was not significant (Figure 6.3C, *P*= 0.41).

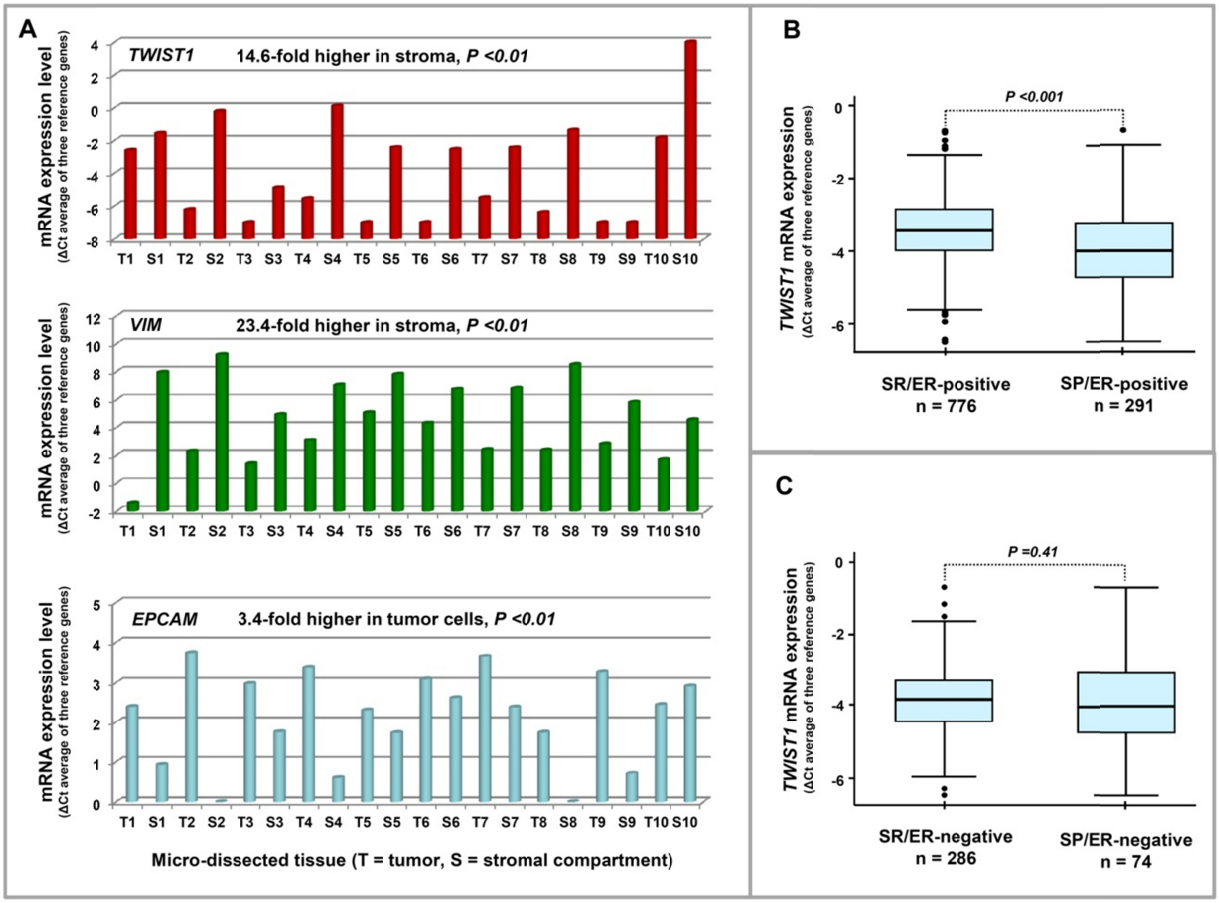


Figure 6.3 Association of *TWIST1* mRNA expression with stromal content of the tumor tissue. **A.** *TWIST1* mRNA expression levels in laser capture microdissected breast tumor tissues. The stromal and epithelial tumor cell areas were separately analyzed after microdissecting 10 breast tumor tissues (n = 9 with high *ESR1*expression, n = 1, (T6), with low *ESR1* expression). After LCM, RNA isolation and cDNA synthesis were done as described in the Patients and Methods. *TWIST1* mRNA levels were measured by qRT-PCR in RNA isolated from both the stromal (S) and the epithelial tumor cell (T) compartments of these tissues, and levels were normalized to our set of reference genes by the delta (Δ) Ct method. To ensure our set of reference genes did not cause a bias in these analyses, and our LCM material was indeed enriched for stromal and epithelial cells, we also measured *VIM* and *EpCAM* in these fractions. A 2-sided paired t-test was performed to evaluate the differences. **B.** *TWIST1* mRNA expression in SR and SP ER-positive tumors and **C.** *TWIST1* mRNA expression in SR- and SP ER-negative tumors. For figures B and C, the mRNA expression of *TWIST1* was log-transformed and the expression relative to reference genes is presented on the y-axis of the box-and-whisker plots. The box-plot shows the five statistics (box bottom-line is the 25th percentile, solid line in the middle of the box presents the median, upper line of the box is the 75th percentile and the whiskers extend to 1.5 times the inter-quartile range, observations beyond these values are plotted, separately). Mann-Whitney test was performed to evaluate the difference between the SR and the SP groups in both ER-positive and ER-negative patients (B, C).

This altogether suggests that *TWIST1* mRNA is predominantly expressed in the stromal compartment of breast tumor tissues. Besides addressing the issue of where *TWIST1* mRNA is predominantly expressed in breast cancer tissues as described above, we also performed IHC staining to ascertain the localization of TWIST1 protein. For this, we stained 20 tumor tissues with different *TWIST1* mRNA expression (n = 10 with high *TWIST1* mRNA expression, n = 10 with low *TWIST1* mRNA expression). Apparently, TWIST1 protein is predominantly localized in the nuclei of both epithelial tumor cells and in almost all non-epithelial cells present in the stromal

compartment of tissues, such as fibroblast cells, endothelial cells, and inflammatory cells (Figure 6.4 A-B).

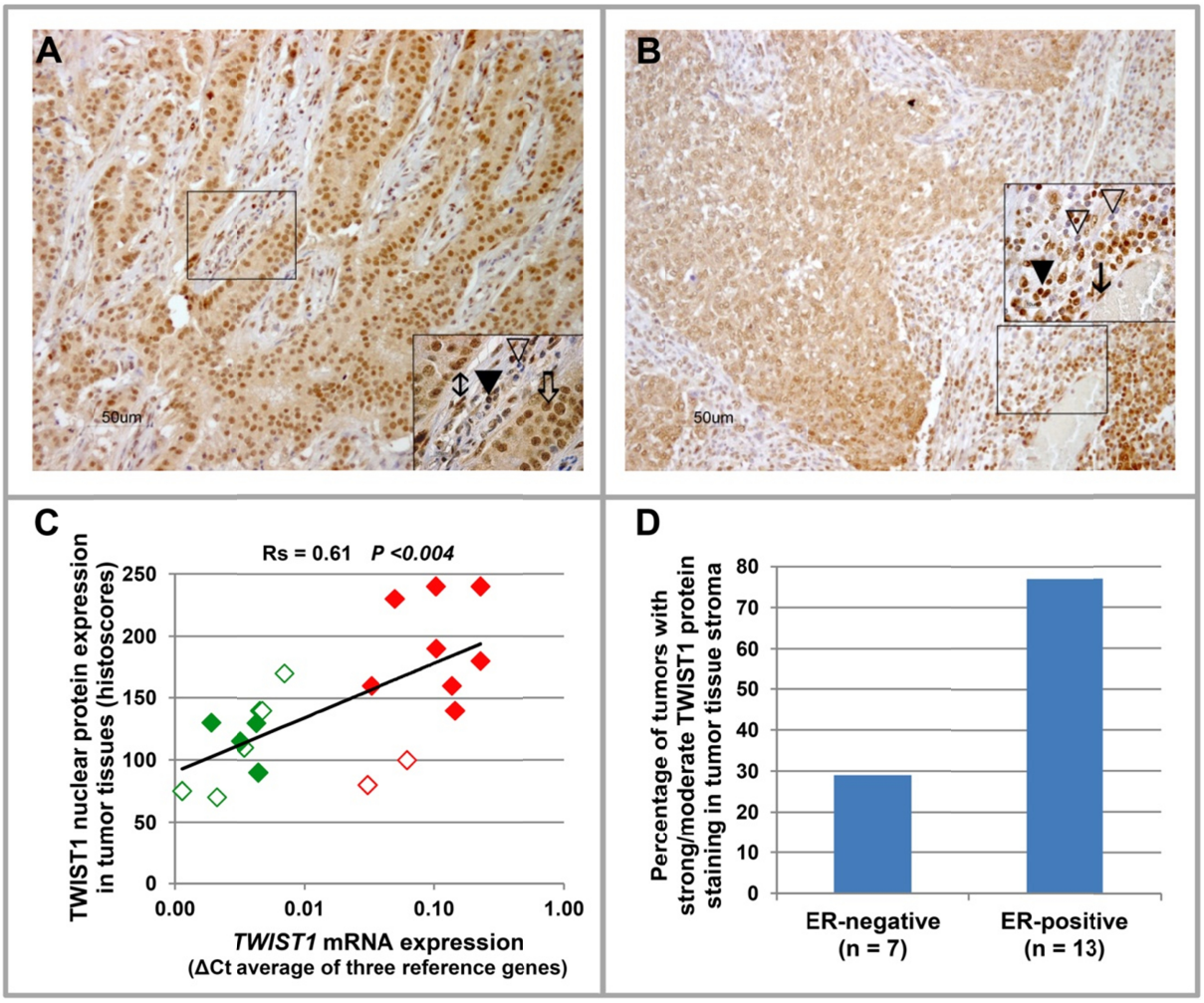


Figure 6.4 Representative tissues with IHC staining of TWIST1 protein. **A.** Strong nuclear staining of TWIST1 protein in an ER-positive tumor tissue with high *TWIST1* mRNA expression. **B.** Weak nuclear staining of TWIST1 protein in an ER-negative tumor tissue with low *TWIST1* mRNA expression. **C.** Correlation of *TWIST1* mRNA expression and protein expression in breast tumor tissues (n = 20; n = 10 high *TWIST1* mRNA expression, n = 10 low *TWIST1* mRNA expression). Red and green boxes represent high and low *TWIST1* mRNA expressing tumor tissues, respectively, whereas empty boxes and filled boxes represent ER-negative and ER-positive tumor tissues, respectively. Spearman rank correlation test was used to evaluate the correlation between *TWIST1* mRNA and protein expression. **D.** Association of stromal staining of TWIST1protein with ER-positive tumor tissues. (▽= negative nuclei of inflammatory cells; ▼= strong positive nuclei of inflammatory cells; ⇕ = strong positive nuclei of fibroblast; ⇨= strong positive nuclei of tumor cells; ↓ = strong positive nuclei of endothelial cells)

In these 20 tumors, *TWIST1* mRNA expression levels were significantly correlated with the nuclear protein expression in tumor tissue (Spearman rank correlation, Rs = 0.61, *P* <0.004) (Figure 6.4C). Additionally, tumor tissues showing weak/moderate TWIST1 protein staining were mostly ER-negative (Figure 6.4C).

Our data suggest that stroma of ER-positive (and not ER-negative tumors) is enriched for TWIST1. To address this in more detail, we stained 20 primary breast tumor tissues (n = 7 ER-negative and n =13 ER-positive) for TWIST1 protein. We observed that 2 out of 7 (29%) of the ER-negative tumor tissues and 10 out of 13 (77%) of the ER-positive tumor tissues showed moderate to strong TWIST1

protein staining in the stroma. The staining of TWIST1 in the other tumors (71% of the ER-negatives and 23% of the ER-positives) was weak. These data based on limited number of samples further support our qRT-PCR based findings at the protein level that the stroma of ER positive tumors is enriched for TWIST1 expression (Figure 6.4D).

DISCUSSION

The bHLH transcription factor TWIST1, which is essential in developmental processes such as gastrulation, has been shown to be oncogenic in various cancers[26-33]. Functional studies pinpointed a pivotal role for this protein in EMT, a fundamental biological process considered to be crucial for metastatic spread in various carcinomas, including breast cancer. This prompted us to analyze mRNA expression of this gene in primary breast tumors from a large retrospective cohort of patients with complete clinical follow-up and address two basic questions. First, can *TWIST1* mRNA expression in breast cancer tissue predict disease progression? Second, as we had Affymetrix U133A gene-chips data available for a subset of samples[10,11,22], can we reveal which biological pathways are co-expressed with *TWIST1* in breast cancer?

Our current study shows that high *TWIST1* mRNA expression is an independent marker of poor outcome in breast cancer patients. As *TWIST1* mRNA expression was also a predictor of poor outcome in patients with LNN disease who did not receive any adjuvant systemic treatment, we may conclude that *TWIST1* mRNA expression is a pure prognostic marker and is associated with the natural course of disease progression. This suggests that tumors with high expression of *TWIST1* are more aggressive, a finding that fits with experimental data where overexpression of *TWIST1* resulted in increased extravasation during the dissemination of tumor cells[7]. Also in line with this is the increased lymph node involvement of breast tumors exhibiting higher *TWIST1* expression as observed by us and others[34]. The first seminal paper on *TWIST1* in breast cancer also showed higher *TWIST1* mRNA expression in infiltrative lobular carcinoma[7], which we also confirmed though the implications if this remain unclear.

What we further unexpectedly observed was that the association of *TWIST1* expression with poor prognosis was confined to ER-positive cases and not seen in ER-negative breast cancer. Similar observations were recently made by Van Nes and colleagues[21]. This group, by studying *TWIST1* expression in a smaller-sized cohort using IHC on tissue microarrays, also found significantly more cumulative relapses in the ER-positive subgroup of tumors. This observation was not significant in their entire cohort, which also included ER-negative cases. Using the same antibody, we observed, in a subset of formalin-fixed specimen available from our cohort, clear nuclear TWIST1 expression. This protein expression was not only confined to the nuclei of invasive tumor cells but also present in nuclei of other cells, such as fat cells, fibroblasts, endothelial cells and inflammatory cells. Furthermore, although based on small numbers, TWIST1 protein staining was high in those specimens in which we had measured high *TWIST1* mRNA expression in a matched fresh-frozen sample of the same tumor, while in specimens with low *TWIST1* mRNA expression the protein staining was overall weaker or absent.

To study the potential biological mechanisms in disease progression connected to *TWIST1* expression, we identified genes co-expressed with *TWIST1* in LNN/ER-positive tumors from which genome wide gene expression data were available. Three related pathways were revealed. The first pathway was "Focal adhesion" which involves interaction of ECM with integrin-A and -B ligands. The second pathway, "ECM-receptor interaction" is related to ECM, and it involves collagen, laminin, fibronetic, tenascin and integrin-A and -B proteins. The third pathway is the "TGF-beta signaling". How these pathways, involving well-known cancer-related genes are connected to *TWIST1* expression in relation to disease progression in LNN/ER-positive cancer warrants functional validation. A common denominator among these three pathways, however, is their involvement in ECM remodeling. There is a growing body of evidence showing that changes in the tumor microenvironment can promote proliferation of the tumor cells and therefore can influence outcome of patients[35], and our data connect this to *TWIST1* expression. Furthermore, our work also shows that the stroma is a significant source of *TWIST* transcript expression. Indeed, the current data reinforces recent observations made by Roman-Perez and colleagues after evaluating gene expression patterns in 72 breast tissue samples[36]. They identified two distinct subtypes, Active and Inactive, in the cancer-adjacent extra-tumoral microenvironment of breast tissues and demonstrated that *TWIST1* was highly expressed along with other stromal associated genes such as *VIM, ADAMTS2, COL4A2, COL4A1, ITGA7*, and *ITGB1* in the Active subtype. This could suggest a potential involvement of TWIST1 in the activation of stroma in malignant breast tissue. Furthermore, the Active subtype was associated with expression of TGF-beta induced fibroblast activation signatures, and showed a strong association with OS among specifically ER-positive patients. Intriguingly, also in our data set, the association of *TWIST1* with the TGF-beta signaling pathway appears to be more prominent in ER-positive tumors.

Another recent study also demonstrated that TWIST1 protein can be found in both compartments of malignant breast tissue, with a predominant expression in the stromal compartment[37]. More importantly, they found that, although rare in their hands, nuclear expression of TWIST1 protein in the epithelial tumor cell compartment to be associated with poor prognosis (OS). TWIST1 protein expression in the stromal compartment was not associated with OS, but positively associated with a positive ER or PGR status of the tumors.

This mutual dependence of *TWIST1* expression on the presence of a notable stromal component and an ER-positive status of the tumor tissue in our and other studies is in line with the observed prognostic value of *TWIST1* in ER-positive disease only. Although our study is a correlative study and cannot reveal causality, it is intriguing and points to a novel role of *TWIST1* in breast cancer.

What is surprising about our results, which associate high *TWIST1* mRNA expression with poor prognosis in ER-positive breast cancer and with ECM and stromal content, is the fact that up until now the process of EMT has been more frequently assigned to ER-negative cancer and was rarely observed in ER-positive breast cancer. Also, previously studied breast cancer cell lines with clear EMT-like features were ER-negative[38]. This left us with an apparent contradiction. We, therefore, specifically explored whether markers of EMT present on gene expression arrays [11] were associated with *TWIST1* expression. Of the eight markers analyzed, we found Zinc finger E-box-binding

homeobox 1 (*ZEB1)*, Zinc finger E-box-binding homeobox 2 *(ZEB2),* Snail2 (*SNAI2*), Smooth muscle actin also known as Alpha-actin-2 (*ACTA2*) and *VIM* clearly positively correlated with *TWIST1* expression (Rs = 0.34, 0.34, 0.35, 0.42, and 0.28 respectively; all *P* values <0.001; Additional file 6.3, Table S15). This does, in line with the recent study of Roman-Perez and colleagues [28] suggest that markers of EMT are co-expressed with *TWIST1*. However, the lack of a clear negative association of *E-cadherin* mRNA (*CDH1*) with *TWIST1* mRNA expression in our data set could indicate that EMT may not be fully established. Alternatively, and more in line with our findings, is that some markers of EMT such as Vimentin and Smooth muscle actin are also expressed in re-activated stroma[39]. Therefore, our findings are a starting point for more functional studies to be certain about TWIST1 possible role in EMT as well as in re-activated stroma in ER-positive breast cancer.

CONCLUSIONS

In summary, we have found that *TWIST1* mRNA expression level is associated with small tumor size, invasive lobular carcinoma, stromal-rich tumors and with *PGR* and *ERBB2* (over)expression. More importantly, *TWIST1* mRNA expression predicts, independently of the traditional prognostic factors, poor prognosis in patients with primary breast cancer. This association with poor prognosis is particularly observed in ER-positive disease. For this large retrospective study that involved 1,427 breast cancer specimens, we made use of all clinical material available that met the preset criteria described in Figure 1, thus reaching Level of Evidence III (LOE-III). To reach LOE-I or II, a prospective randomized study or pooled meta-analysis is required. Studying *TWIST1* co-expressed genes, we found that ECM related pathways as well as stromal content are correlated with *TWIST1* mRNA expression in LNN/ER-positive breast cancer. Functional studies are therefore required to determine the causal relation between these pathways and the prognostic value of *TWIST1* mRNA expression in ER-positive tumors.

ABBREVIATIONS

ACTA2: alpha-actin-2; BM2: beta-2-microglobulin; CI: confidence interval; ECM: extracellular matrix; EMT: epithelial to mesenchymal transition; EpCAM: epithelial cell adhesion molecule; ER: estrogen receptor; PGR: progesterone receptor; ERBB2: erythroblastic leukemia viral oncogene homolog 2; ESR1: estrogen receptor 1; GEO: gene expression omnibus; GO: gene ontology; H&E: hematoxylin and eosin; HER2: human epidermal growth factor receptor 2; HMBS: hydroxymethylbilane synthase; HPRT1: hypoxanthine phosphoribosyltransferase 1; HR: hazard ratio; IDC: invasive ductal carcinoma; IHC: immunohistochemistry; ILC: invasive lobular carcinoma; KEGG: kyoto encyclopedia of genes and genomes; LCM: laser capture microdissection; LNN: lymph node-negative; MFS: metastasis-free survival; OS: overall survival; PEN: polyethylene naphthalate; Q: quartile; RT-PCR: reverse transcriptase polymerase chain reaction; SP: stromal poor; SR: stromal rich; TWIST1: Twist-related protein 1, also known as a basic helix-loop-helix protein 38 (bHLHa38); VIM: vimentin; ZEB1: zinc finger e-box-binding homeobox 1; ZEB2: zinc finger e-box-binding homeobox 2.

CONFLICT OF INTEREST STATEMENT

All authors declared that they have no competing interest

AUTHORS' CONTRIBUTION

AMS, JAF and JWMM designed the study. MR, JAF and JWMM wrote the manuscript. AMS performed the mRNA expression studies. MAT performed the immunohistochemistry. MR, AMS, JAF and JWMM analyzed the data. MR, MPL and MS did the statistical data analyses. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENT

We are grateful to Joan Bolt, Vanja de Weerd, Anna van Galen and NingQing Liu for their contribution and technical support. We appreciate Gaia Schiavon for her thoughtful discussion. This study was in part financed by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO). MR was funded by the Higher Education Commission (HEC) government of Pakistan.

ADDITIONAL FILES

Additional file 6.1
Figure S1 and S2

Additional file 6.2
Tables S1-8

Additional file 6.3
Tables S.9-15

Additional files are available on:
https://www.dropbox.com/sh/4fwrk0lit91dsp5/WP1sZYZAQw?m

REFERENCES

1. Jemal, A. et al. Global cancer statistics. *CA Cancer J Clin* **61**, 69-90 (2011).
2. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).

3. Firulli, A.B. & Conway, S.J. Phosphoregulation of Twist1 provides a mechanism of cell fate control. *Curr Med Chem* **15**, 2641-7 (2008).
4. Thisse, B., el Messal, M. & Perrin-Schmitt, F. The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Res* **15**, 3439-53 (1987).
5. Leptin, M. twist and snail as positive and negative regulators during Drosophila mesoderm development. *Genes Dev* **5**, 1568-76 (1991).
6. Kress, W. et al. Saethre-Chotzen syndrome caused by TWIST 1 gene mutations: functional differentiation from Muenke coronal synostosis syndrome. *Eur J Hum Genet* **14**, 39-48 (2006).
7. Yang, J. et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-39 (2004).
8. Yang, M.H. et al. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nature cell biology* **12**, 982-92 (2010).
9. Mani, S.A. et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-15 (2008).
10. Yu, J.X. et al. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer* **7**, 182 (2007).
11. Wang, Y. et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671-9 (2005).
12. Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands [http://www.federa.org/?s=1&m=99].
13. McShane, L.M. et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast cancer research and treatment* **100**, 229-35 (2006).
14. Altman, D.G., McShane, L.M., Sauerbrei, W. & Taube, S.E. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* **10**, 51 (2012).
15. Sieuwerts, A.M. et al. How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 7311-21 (2005).
16. Sieuwerts, A.M. et al. Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer. *Clinical chemistry* **53**, 1280-8 (2007).
17. van Agthoven, T. et al. Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J Clin Oncol* **27**, 542-9 (2009).
18. Foekens, J.A. et al. The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer research* **60**, 636-43 (2000).
19. Sieuwerts, A.M. et al. Which cyclin E prevails as prognostic marker for breast cancer? Results from a retrospective study involving 635 lymph node-negative breast cancer patients.

Clinical cancer research : an official journal of the American Association for Cancer Research **12**, 3319-28 (2006).

20. Umar, A., Dalebout, J.C., Timmermans, A.M., Foekens, J.A. & Luider, T.M. Method optimisation for peptide profiling of microdissected breast carcinoma tissue by matrix-assisted laser desorption/ionisation-time of flight and matrix-assisted laser desorption/ionisation-time of flight/time of flight-mass spectrometry. *Proteomics* **5**, 2680-8 (2005).
21. van Nes, J.G. et al. Co-expression of SNAIL and TWIST determines prognosis in estrogen receptor-positive early breast cancer patients. *Breast cancer research and treatment* (2011).
22. Smid, M. et al. Subtypes of breast cancer show preferential site of relapse. *Cancer research* **68**, 3108-14 (2008).
23. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
24. Tong, W. et al. ArrayTrack--supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ Health Perspect* **111**, 1819-26 (2003).
25. ArrayTrack™—a microarray database, data analysis & interpretation tool [http://www.fda.gov/ScienceResearch/BioinformaticsTools/Arraytrack/default.htm].
26. Valsesia-Wittmann, S. et al. Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer cell* **6**, 625-30 (2004).
27. Fondrevelle, M.E. et al. The expression of Twist has an impact on survival in human bladder cancer and is influenced by the smoking status. *Urol Oncol* **27**, 268-76 (2009).
28. Elias, M.C. et al. TWIST is expressed in human gliomas and promotes invasion. *Neoplasia* **7**, 824-37 (2005).
29. Ru, G.Q., Wang, H.J., Xu, W.J. & Zhao, Z.S. Upregulation of Twist in gastric carcinoma associated with tumor invasion and poor prognosis. *Pathol Oncol Res* **17**, 341-7 (2011).
30. Lee, K.W. et al. Twist1 Is an Independent Prognostic Factor of Esophageal Squamous Cell Carcinoma and Associated with Its Epithelial-Mesenchymal Transition. *Ann Surg Oncol* (2011).
31. Okada, T. et al. TWIST1 hypermethylation is observed frequently in colorectal tumors and its overexpression is associated with unfavorable outcomes in patients with colorectal cancer. *Genes Chromosomes Cancer* **49**, 452-62 (2010).
32. Gomez, I. et al. TWIST1 is expressed in colorectal carcinomas and predicts patient survival. *PloS one* **6**, e18023 (2011).
33. Cosset, E. et al. Deregulation of TWIST-1 in the CD34+ compartment represents a novel prognostic factor in chronic myeloid leukemia. *Blood* **117**, 1673-6 (2011).
34. Valdes-Mora, F. et al. TWIST1 overexpression is associated with nodal invasion and male sex in primary colorectal cancer. *Ann Surg Oncol* **16**, 78-87 (2009).

35. Bergamaschi, A. et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J Pathol* **214**, 357-67 (2008).
36. Roman-Perez, E. et al. Gene expression in extratumoral microenvironment predicts clinical outcome in breast cancer patients. *Breast Cancer Res* **14**, R51 (2012).
37. Soini, Y. et al. Transcription factors zeb1, twist and snai1 in breast carcinoma. *BMC Cancer* **11**, 73 (2011).
38. Blick, T. et al. Epithelial mesenchymal transition traits in human breast cancer cell lines. *Clin Exp Metastasis* **25**, 629-42 (2008).
39. Farmer, P. et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* **15**, 68-74 (2009).

Chapter 7

MicroRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs

Muhammad Riaz[*]**,** Marijn T.M. van Jaarsveld[*], Antoinette Hollestelle, Wendy J.C. Prager-van der Smissen, Anouk A.A.J. Heine, Antonius W.M. Boersma, Jingjing Liu, Jean Helmijr, Bahar Ozturk, Marcel Smid, Erik A. Wiemer, John A. Foekens, and John W.M. Martens.

Submitted for publication

ABSTRACT

Introduction Breast cancer is a genetically and phenotypically complex disease. To understand the role of microRNAs (miRNAs) in this molecular complexity, we performed miRNA expression analysis in a cohort of molecularly well-characterized human breast cancer (BC) cell lines to identify miRNAs associated with the most common molecular subtypes and the most frequent genetic aberrations.

Methods Using a microarray carrying LNA™ modified oligonucleotide capture probes (Exiqon), expression levels of 725 human miRNAs were measured in 51 BC cell lines. MiRNA expression was explored by unsupervised cluster analysis and then associated with the molecular subtypes and genetic aberrations commonly present in breast cancer.

Results Unsupervised cluster analysis using the most variably expressed miRNAs divided the 51 BC cell lines into a major and a minor cluster predominantly mirroring the luminal and basal intrinsic subdivision of BC cell lines. One hundred and thirteen miRNAs were differentially expressed between these two main clusters. Forty miRNAs were differentially expressed between basal-like and normal-like/claudin-low cell lines. Within the luminal-group, 39 miRNAs were associated with *ERBB2* overexpression and 24 with *E-cadherin* gene mutations, which are frequent in this subtype of BC cell lines. In contrast, 31 miRNAs were associated with *E-cadherin* promoter hypermethylation, which, contrary to *E-cadherin* mutation, is exclusively observed in BC cell lines that are not of luminal origin. Thirty miRNAs were associated with $p16^{INK4}$ status while only a few miRNAs were associated with *BRCA1*, or *PIK3CA*/*PTEN, TP53* mutation status. Twelve miRNAs were associated with DNA copy number variation of the respective locus.

Conclusion Luminal-basal and epithelial-mesenchymal associated miRNAs determine the subdivision of miRNA transcriptome of BC cell lines. Specific sets of miRNAs were associated with *ERBB2* overexpression, $p16^{INK4a}$or *E-cadherin* mutation or *E-cadherin* methylation status, which implies that these miRNAs may contribute to the driver role of these genetic aberrations. Additionally, miRNAs, which are located in a genomic region showing recurrent genetic aberrations, may themselves play a driver role in breast carcinogenesis or contribute to a driver gene in their vicinity. In short, our study provides detailed molecular miRNA portraits of BC cell lines, which can be exploited for functional studies of clinically important miRNAs.

INTRODUCTION

It is becoming apparent that breast cancer (BC) is genetically and epigenetically not just one disease but a diverse group of diseases with diverse clinically relevant biological and phenotypical features. Recent technological advances in molecular profiling have led to the identification of an increasing number of molecular subtypes in breast cancer, each with distinct co- and anti-regulated genes. However, the biology of these molecular subtypes and their underlying genetic drivers may be affected by numerous biological factors including miRNAs. MiRNAs are a class of small non-protein-coding genes, which regulate the expression of genes post-transcriptionally via sequence-specific interaction with the 3' untranslated region (UTR) of target mRNAs resulting in inhibition of translation and/or mRNA degradation[1,2]. A large number of studies have established that miRNAs play essential roles in biological processes, such as development[3,4], cell proliferation [5], apoptosis [6], stress response, and tumorigenesis [7,8]. Aberrant expression levels of miRNAs have been observed in many solid cancers including breast cancer. In breast cancer, the expression levels of several miRNAs are significantly different between normal and cancerous tissues, between breast cancers of different molecular subtypes[9-11], with a different prognosis[12-14], and with breast cancers showing different response to endocrine therapy[15,16]. Despite the significant progress in the last few years, the exact biological functions and the genetic factors driving their expression have been revealed for only a limited number of breast cancer miRNAs.

Human BC cell lines are excellent experimental models and renewable resources to investigate biological functions of clinically important miRNAs both in *in vitro* cultured conditions and *in vivo* when raised as xenografts[6,17-20]. Here, using microarrays we analyzed miRNA expression levels in 51 molecularly well-characterized human BC cell lines. We explored the association of individual miRNA expression levels with intrinsic subtypes and the most common recurrent genetic aberrations. By doing this, we provide a catalog of miRNAs in human BC cell lines, which can be used to understand underlying biology of clinically relevant miRNAs and to reveal the genetic factors that may be involved in their regulation.

MATERIALS AND METHODS

Breast cancer cell lines

The cohort of 51 human breast cancer cell lines used in this study is listed in Table 7.1. The origin of the cell lines has been described elsewhere[21]. Prior to expression profiling, all cell lines were established to be genetically unique, monoclonal and of correct identity by performing short tandem repeat (STR) profiling using the PowerPlex® 16 System (Promega, Madison, WI, USA). The PowerPlex® 16 system included fifteen STRs and one gender discriminating locus (Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317, and D5S818). The method involved isolation of the genomic DNA from each BC cell line using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) and 10 ng of the isolated DNA was used as input for the multiplex PCR. The following PCR conditions were used: 95°C for 11 min, 96°C

for 1 min, 10x (94°C for 30 sec, 60°C for 30 sec, 70°C for 45 sec), 22x (90°C for 30 sec, 60°C for 30 sec, 70°C for 45 sec), and then 60°C for 30 min. The PCR was carried out using primers linked with fluorescent dyes (6-carboxy-4', 5'-dichloro-2', 7'-dimethoxy-fluorescein, Fluorescein, and Carboxy-tetramethylrhodamine). The labeled amplicons were detected using the 3130*xl* Genetic Analyzer (Applied Biosystems, USA) and data was analyzed using the Genemarker 1.91 software from Softgenetics. The end result for each cell line was an electropherogram with each STR allele represented as one or more peaks of an appropriate fluorophore. The authenticity of all cell lines except SUM cell lines, were assessed by comparing the generated STR profiles with the source STR profiles present in the American Type Culture Collection (ATCC) and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). As no reference is available, the STR profiles of the SUM cell lines were matched with the profiles generated from the earliest passage of these cell lines stored in the in-house culture collection.

For experiments, each cell line was cultured in triplicate on collagen-coated petri dishes in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum and antibiotic agents (100 µg/ml penicillin G and 80 µg/m streptomycin (P/S). The petri dishes were placed in a humidified atmosphere of 5% CO_2 and 95% air at 37°C until cultures were 70-80% confluent.

Table 7.1 Molecular and biochemical characterization of 51 human breast cancer cell lines

Cell lines	Molecular classification	Clinically relevant protein expression							
	Intrinsic subtype[$]	ER	PGR	ERBB2	EGFR	CK5	CK8-18	CK19	CK14
SUM185PE	Luminal	–	–	–	–	–	+	+	–
MDA-MB-175VII	Luminal	+	–	–	–	–	+	+	–
BT483	Luminal	+	–	–	–	–	+	+	–
T47D	Luminal	+	+	–	+	–	+	+	–
MDA-MB-415	Luminal	+	+	–	–	–	+	+	–
ZR75-1	Luminal	+	+	–	–	–	+	+	–
MCF-7	Luminal	+	+	–	–	–	+	nd	–
SUM52PE	Luminal	+	–	–	–	–	+	nd	–
UACC812	Luminal	+	+	++	–	–	+	+	–
MDA-MB-361	Luminal	+	+	++	–	–	+	+	–
ZR75-30	Luminal	+	–	++	–	–	+	+	–
SK-BR-5	Luminal	–	–	++	–	–	+	+	–
OCUB-F	Luminal	–	–	++	–	+	+	+	–
MPE600	Luminal	+	–	++	–	–	+	+	–
MDA-MB-134VI	Luminal	+	–	–	–	–	+	+	–
SUM44PE	Luminal	+	+	–	–	–	+	+	–
CAMA-1	Luminal	+	–	–	–	–	+	+	–
BT474	Luminal	+	+	++	–	–	+	+	–
MDA-MB-330	Luminal -ERBB2 +	+	–	++	–	–	+	+	–
HCC1419	Luminal -ERBB2 +	+	–	++	–	–	+	+	–
HCC202	Luminal -ERBB2 +	–	–	++	–	–	+	+	–
SUM190PT	Luminal -ERBB2 +	–	–	++	–	–	+	+	–
SUM225CWN	Luminal -ERBB2 +	–	–	++	–	–	+	+	–
UACC893	Luminal -ERBB2 +	–	–	++	+	–	+	+	–
SK-BR-3	Luminal -ERBB2 +	–	–	++	–	–	+	+	–
EVSA-T	Luminal -ERBB2 +	–	+	++	–	–	+	+	–
MDA-MB-453	Luminal -ERBB2 +	–	–	++	–	–	+	+	–
HCC1569	ER-negative-ERBB2 +	–	–	++	+	–	–	–	–
HCC1954	ER-negative-ERBB2 +	–	–	++	+	+	+	–	–

Table 7.1 Continued

Cell lines	Molecular classification	Clinically relevant protein expression							
	Intrinsic subtype[$]	ER	PGR	ERBB2	EGFR	CK5	CK8-18	CK19	CK14
HCC1500	ER-negative-ERBB2 +	+	+	–	–	–	+	+	–
DU4475	Other	–	–	–	–	+	–	–	–
SUM229PE	Basal-like	–	–	–	+	+	+	+	–
HCC1937	Basal-like	–	–	–	+	+	+	–	–
MDA-MB-468	Basal-like	–	–	–	+	+	+	+	–
HCC1806	Basal-like	–	–	–	+	+	+	+	+
HCC70	Basal-like	–	–	–	+	+	+	+	–
HCC1143	Basal-like	–	–	–	+	+	+	+	–
BT20	Basal-like	–	–	–	+	+	+	+	–
SUM149PT	Basal-like	–	–	–	+	+	+	+	–
HCC1395	Basal-like	–	–	–	+	+	–	–	–
SK-BR-7	Normal-like/claudin-low	–	–	–	+	–	+	–	–
Hs578T	Normal-like/claudin-low	–	–	–	+	–	–	–	–
MDA-MB-231	Normal-like/claudin-low	–	–	–	+	–	–	–	–
SUM1315M02	Normal-like/claudin-low	–	–	–	+	–	–	–	–
MDA-MB-436	Normal-like/claudin-low	–	–	–	+	–	–	–	–
BT549	Normal-like/claudin-low	–	–	–	+	–	–	–	–
MDA-MB-157	Normal-like/claudin-low	–	–	–	+	–	–	–	–
SUM159PT	Normal-like/claudin-low	–	–	–	+	–	–	–	–
MDA-MB-435s	Normal-like/claudin-low	–	–	–	–	–	–	–	–
SUM102PT	Normal-like/claudin-low	nd	nd	nd	nd	nd	nd	nd	nd
HCC38	Normal-like/claudin-low	nd	nd	nd	nd	nd	nd	nd	–

$ Cell lines are classified based on expression of the intrinsic gene set defined by Perou and colleagues [23]
For protein expression: +, expression; ++, overexpression; –, no expression; and nd, not determined.
Other subtype includes DU4475 since it is not classified to any subtype of breast cancer by intrinsic gene set.

Total RNA isolation

Total RNA from all samples was isolated using RNAzol-B reagent (Campro Scientific BV, Veenendaal, The Netherlands) according to the manufacturer's manual. Briefly, a biological sample was lysed in RNAzol-B reagent and the lysate was separated into an aqueous and an organic phase after the addition of chloroform. DNA and protein were subsequently removed by carefully transferring the aqueous phase containing RNA to a fresh eppendorf. The RNA was obtained from the aqueous phase by an isopropanol precipitation, washed with ethanol and air dried for subsequent procedures. The purity of the isolated RNA was checked using NanoDrop® ND-1000 ensuring spectrophotometric ratios of A260nm/A280nm ~ 2 and A260nm/A230nm ≥ 2 and the quality control checks were performed according to the previously described methodology[22].

Gene expression profiling

Two hundred ng of total RNA was reverse transcribed, copied into double strand cDNA and labeled to yield biotin-labeled cRNA using the 3′ IVT express Kit (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's instructions. Biotin labeled cRNA was subsequently fragmented and loaded onto an Affymetrix GeneTitan instrument. The hybridization cocktail was applied to Human Genome HT_HG-U133_Plus_PM GeneChip 96-well arrays. All steps including hybridization, washing and scanning were carried out automatically inside the instrument. The raw data (.CEL files) were normalized by RMA method using the default settings of the Affymetrix® Expression

Console™ software and used for statistical analysis. The microarray data has been deposited to Gene Expression Omnibus (GEO) data repository under the accession number GSE41313.
For subtype classification of the cell lines, Perou's intrinsic gene set of 496 genes[23] were matched to the Affymetrix probe sets using Unigene cluster numbers. Some of the genes have multiple probe sets present. To ensure analysis of only the informative probe sets, the probe sets that did not vary across all samples were removed, leaving the most variable ones for analysis (66% of the probe sets). These genes were then used to cluster the BC cell lines. The intrinsic molecular subtypes were assigned as follows: luminal-type cell lines that exhibited higher expression of *ESR1, GATA3, TFF3, FOXA1*; ERBB2-positive cell lines that showed higher expression of *ERBB2, GRB7, STARD3*; basal-like cell lines, which were characterized by higher expression of *KRT5, KRT17, BST2, FABP7*; and normal-like/claudin-low cell lines that did not show *KRT5, KRT17* expression and have low expression of claudin 3, 4, and 7 genes. The ERBB2-positive cell lines were further designated as luminal-ERBB2-positive and ER-negative/ERBB2-positive cell lines as the former in addition to *ERBB2* overexpression also show ESR1 gene expression on microarray while the latter do not.

miRNA expression analysis

The miRNA expression profiling was performed using miRNA microarrays according to previously published method[24]. In brief, 1 µg of total RNA was labeled with Cy3 using the ULS aRNA labeling kit (Kreatech, Amsterdam, The Netherlands). The LNA™ modified oligonucleotide capture probe set (Exiqon Vedbaek, Denmark, miRBase version 10.0, annotation version 13) was spotted in duplicate on Nexterion E glass slides in Nexterion Spot buffer using a Virtek Chipwriter Pro. The RNA sample with a labeling efficiency over 15 pmol Cy3/µg RNA was used for hybridization in a salt-based hybridization buffer (Ocimum Biosolutions) overnight at 60°C using a Tecan HS4800 pro hybridization station. Hybridized slides were scanned in a Tecan LS Reloaded scanner. Data was extracted using Imagene software (6.0 standard edition). The raw data were normalized using quantile normalization and used for statistical analysis. The normalized expression data is provided in Additional file 7.1: Table S1.

SNP arrays

Genomic DNA from all cell lines was extracted using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Five hundred ng of genomic DNA was used as starting material to capture genome-wide chromosomal aberrations with the aid of the Affymetrix Genome-Wide Human SNP 6.0 array technology. The steps were performed according to Affymetrix's recommended protocols (Affymetrix, Santa Clara, CA, USA). In summary, after digestion of the genomic DNA using either restriction enzymes NspI or StyI, adaptors were ligated to the obtained DNA fragments. These fragments were subsequently amplified using PCR, fragmented, end-labeled with biotin and hybridized onto GeneChip SNP 6.0 arrays. After hybridization the arrays were washed and scanned to generate the raw data (.CEL files) using the Affymetrix® Genotyping Console™ Software. The chromosomal gains and losses were calculated using SNP copy numbers variation on the same chromosome. The chromosomal regions containing gains or losses were correlated with the

expression level of the miRNAs located on the same genomic regions. The SNP data have been deposited to GEO data repository under the accession number GSE41313.

Protein expression and mutational analysis

Protein expression data of ER, PGR, ERBB2, EGFR, and cytokeratin (CK) 5, CK8-18, CK19, and CK14 were used from previously published work[25] except for 10 HCC cell lines, which were characterized by immunohistochemistry using the same protocols as described before[25]. Mutational analysis of *p16^{INK4a}*, *BRCA1*, *E-cadherin*, *PIK3CA*, and *PTEN* genes and promoter hypermethylation analysis of *E-cadherin* gene were previously reported for all cell lines except for HCC cell lines. These were separately analyzed according to the previously published methods[25] and the data is provided in Additional file7.1: Table S2.

Hierarchical clustering and statistical analyses

Hierarchical clustering analyses of significant miRNAs were performed using Cluster 3 software[26] and the expression patterns of miRNAs and mRNAs in the heat maps were visualized using Treeview 1.1.6 R2. Average linkage clustering was carried out on both samples and mRNA and miRNA expression data respectively, using Pearson correlation as a distance measure. Differential expression of miRNAs between two groups was determined using univariate t-tests in BRB-array tools 3.7. A permutation P-value <0.05 was considered to be statistically significant and used to select differentially expressed miRNAs for supervised hierarchical cluster analysis. The associations between continuous variables were tested using Spearman Rank correlation (Rs). Kruskal-Wallis test was used to evaluate differences among groups and a P-value <0.05 was considered to be statistically significant.

RESULTS

Molecular features of human breast cancer cell lines

The 51 human BC cell lines that were used in this study are listed in Table 7.1 together with the protein expression results of ER, PGR, ERBB2, EGFR and CK5, 8-18, 19, and 14. All cell lines were first profiled for global mRNA expression using microarrays and the cell lines in Table 1 are grouped into molecular subtypes based on the expression of the intrinsic genes originally described by Perou and colleagues[23]. A Pearson correlation based on the top 10% variably expressed genes classified the cell line cohort into two obvious groups (Figure 1). The first major group included 27 cell lines of which 17 expressed ER protein. All cell lines in this group showed higher expression of luminal intrinsic genes as defined by Perou and colleagues[23]. We therefore defined this group of BC cell lines as "luminal-group" (Figure 7.1, left block). On the other hand, the second group included 23 BC cell lines, all of them were ER-negative and showed a predominant expression of basal intrinsic genes. We defined this group of BC cell lines as "ER-negative/basal-group" (Figure 7.1, right block). Moreover, according to classification using the intrinsic gene expressions[23], the luminal-group included 9 ERBB2 overexpressing BC cell lines, which did not cluster as a separate

group. This is in line with clustering of clinical specimens where the majority of luminal, ERBB2-positive tumors cluster with the luminal B tumors[27]. Within the ER-negative/basal-group, the basal-like and normal-like/claudin-low cell lines clustered distinctively into two subgroups (Figure 7.1). The DU4475 cell line could not be assigned to any subtype using intrinsic gene set and was therefore designated as 'other' subtype.

Notably, the entire luminal-group of 27 cell lines expressed luminal cytokeratins (such as CK8/18 and CK19). The basal-like cell lines within ER-negative/basal-group, expressed basal and luminal cytokeratins (CK5, CK19, CK8-18), whereas all normal-like/claudin-low cell lines lacked the expression of these cytokeratins except for SK-BR-7 that exhibited CK8-18 expression (Table 7.1). Furthermore, except for MDA-MB-435s, the entire ER-negative/basal-group of 23 cell lines also expressed EGFR protein (Table 7.1), which is a known marker for basal-type breast cancer.

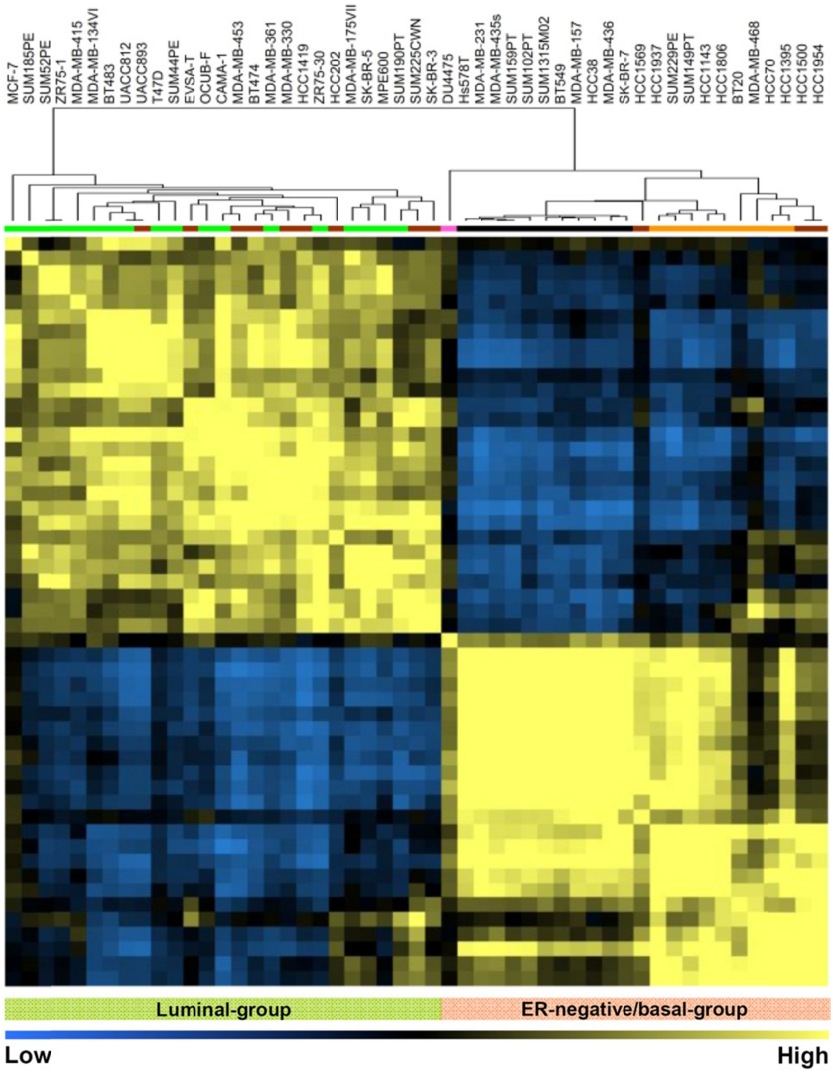


Figure 7.1 Molecular subtyping of 51 human breast cancer cell lines. Pearson correlation plot based on global mRNA expression of the top 10% variable genes. BC cell lines are depicted according to their overall similarity in gene expression. The yellow and blue color indicates high and low overall similarity of samples in mRNA expression, respectively. Two main groups of 27 and 23 cell lines were apparent. Color codes for breast cancer subtypes based on intrinsic gene set are: green, luminal-type cell lines; brown, luminal ERBB2-positive cell lines; black, normal-like/claudin-low cell lines; orange, basal-like cell lines; blue, ER-negative/ERBB2-positive cell lines; pink, other subtype cell lines.

MicroRNA expression profiling revealed two groups of breast cancer cell lines

MiRNA expression profiling was carried out using microarrays containing LNA capture probe sets against 725 human miRNAs. After filtering out miRNAs that were undetectable in most of the cell lines (57%), 413 miRNAs remained and were selected for further analysis. Unsupervised hierarchical clustering based on the most variable 87 miRNAs (with standard deviation $\geq$ 0.75) revealed two prominent cell line clusters; a minor cluster that included 18 and a major cluster that included 33 cell lines (Figure 2). Based on intrinsic subtyping, 17 of 18 cell lines present in the minor cluster belonged to either the basal-like or the normal-like/claudin-low subtypes, and interestingly 16 of them were deficient of ER, PGR and ERBB2 protein expression (triple-negative).

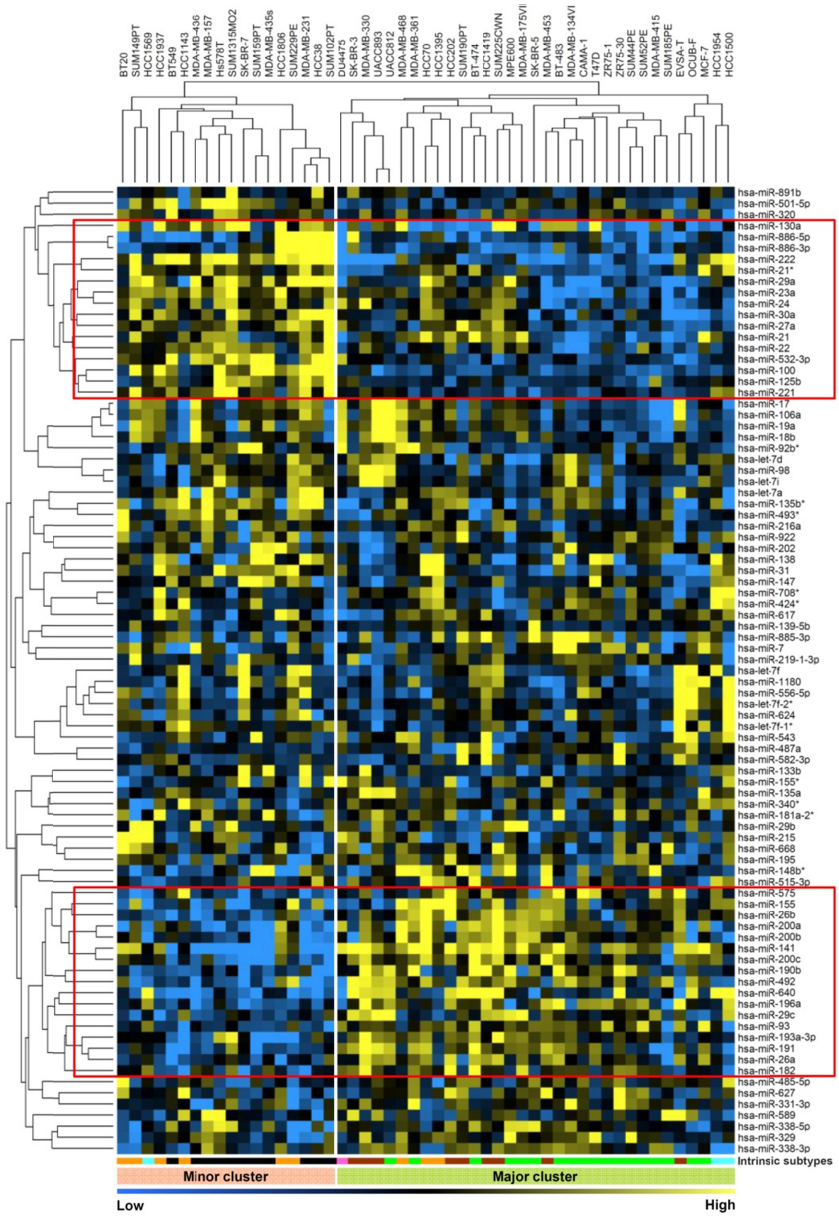


Figure 7.2 Global miRNAs expression analysis of 51 human breast cancer cell lines. Unsupervised clustering across the BC cell lines using the 87 most variably expressed miRNAs. The yellow color indicates high expression and blue low expression of the particular miRNA in the particular sample. The cell lines in the dendrogram of the hierarchical clustering based on miRNAs expression are coded according to their intrinsic subtypes and depicted at the bottom of the figure with the color codes used in Figure 1. Red boxes enlist the groups of miRNAs that show the most robust cluster-specific differential expression in the cell lines. The *hsa-miR-21* miRNA in the red box of minor cluster failed to reach our criteria of fold-change of $\geq$ 1.5.

The other two cell lines (HCC38 and SUN102PT) remained undetermined for protein analysis due to technical reasons. On the other hand 29 of 33 BC cell lines present in the major cluster were of the luminal-type, and all cell lines except SUM185 exhibited either one of the hormone receptor proteins or showed ERBB2 protein expression. The DU4475 cell line, which was previously proposed to have a stem cell-like phenotype due to activation of the Wnt signaling pathway[25,28], clustered at the edge of the major cluster. As a whole, similar to mRNA expression, miRNA expression profiling discriminated cell lines into two major subtypes of breast cancer.

Among the 87 most-variable expressed miRNAs across the entire panel, a group of 15 miRNAs (*hsa-miR-130a, hsa-miR-886-5p, hsa-miR-886-3p, hsa-miR-222, hsa-miR-21*, hsa-miR-29a, hsa-miR-23a, hsa-miR-24, hsa-miR-30a, hsa-miR-27a, hsa-miR-22, hsa-miR-532-3p, hsa-miR-100, hsa-miR-125b, hsa-miR-221*) was significantly higher expressed in the minor cluster as opposed to other miRNAs (Figure 7.2, top red box). Another group of 17 miRNAs (*hsa-miR-575, hsa-miR-155, hsa-miR-26b, hsa-miR-200a, hsa-miR-200b, hsa-miR-141, hsa-miR-200c, hsa-miR-190b, hsa-miR-492, hsa-miR-640, hsa-miR-196a, hsa-miR-29c, hsa-miR-93, hsa-miR-193a-3p, hsa-miR-191, hsa-miR-26a, hsa-miR-182*) showed significantly higher expression in the major cluster compared to the other miRNAs (Fold change $\geq$ 1.5) (Figure 7.2, bottom red box).

As the miRNA-driven major and minor clusters included ER-positive (18 of 33) and ER-negative (all) cell lines, respectively, we determined whether differential expression of miRNAs between the two clusters is driven by ER status of the cell lines. For this, we grouped cell lines into ER-positive and ER-negative based on mRNA expression of ER and performed supervised cluster analysis on these groups. We identified 79 differentially expressed miRNAs between ER-positive and ER-negative cell lines (*P <0.05*) (Additional file 7.2: Figure S1 and Additional file 7.1: Table S3), of which 54 miRNAs were common to the 113 miRNAs that were differentially expressed between the two clusters identified through unsupervised analysis (Additional file 7.1: Table S4). Interestingly, however, 59 miRNAs were uniquely differentially expressed between the two miRNA-driven clusters whereas 25 miRNAs were uniquely related to ER status of the cell lines (Additional file 7.1: Table S5).

MiRNAs expression of cell lines associated with molecular intrinsic subtypes of breast cancer

Next, we aimed to identify miRNAs, which are differentially expressed between the intrinsic subtypes of breast cancer. As expected, we observed a substantial overlap of differentially expressed miRNAs between the luminal-group and ER-negative/basal-group of cell lines and those between ER-positive and ER-negative cell lines (Additional file2: Figure S1, Additional file7.1: Tables S3 and S6). To identify additional miRNAs, which are related to intrinsic subtypes but to avoid the confounding effect of ER, we compared cell lines with and without ERBB2 overexpression within the luminal-group. Likewise we compared the intrinsic basal-like cell lines with normal-like/claudin-low cell lines within the ER-negative/basal-group. These comparisons revealed 39 differentially expressed miRNAs in luminal cell lines with or without overexpressed ERBB2; 40 miRNAs were found differentially expressed between basal-like and normal-like/claudin-low cell lines (*P <0.05*) (Figure 7.3A and B; Additional file 7.1: Tables S7 and S8).

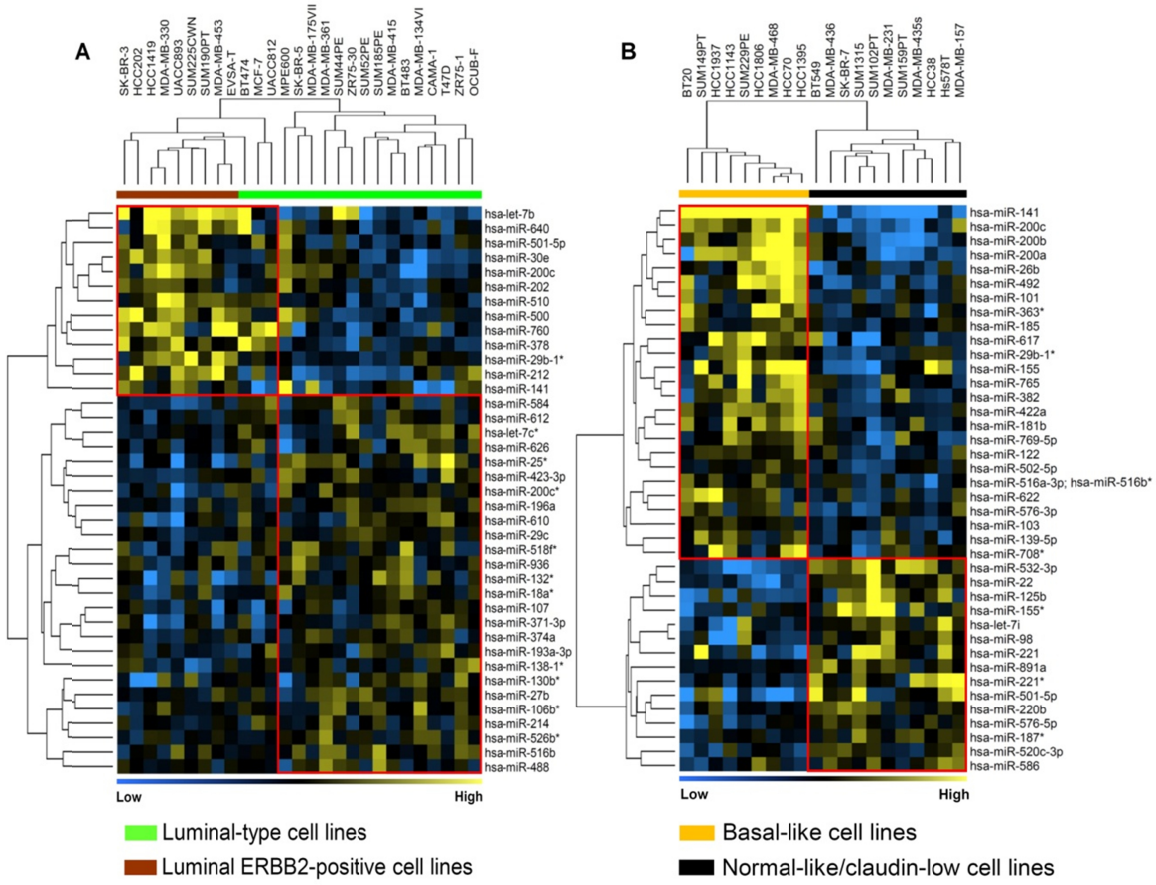


Figure 7.3 Molecular subtype-specific differential expression of miRNAs. **A.** Differentially expressed miRNAs between luminal-type and luminal ERBB2-positive BC cell lines within mRNA derived luminal-group (see Figure 1) and **B.** between basal-like and normal-like/claudin-low BC cell lines within mRNA derived ER-negative/basal-group. The yellow and blue color indicates respectively high and low overall similarity of samples in miRNA expression.

The top most highly expressed miRNAs in ERBB2 overexpressing cell lines included *hsa-let-7b*, *hsa-miR-640*, *hsa-miR-200c*, *hsa-miR-378*, *hsa-miR-141*, *hsa-miR-196a*, *hsa-miR-29c*, and *hsa-miR-18a** whereas *hsa-miR-501-5p*, *hsa-miR-202*, *hsa-miR-760*, and *hsa-miR-626* were more highly expressed in luminal cell lines lacking ERBB2 over-expression (Fold change $\geq$ 1.5) (Additional file 7.1: Table S7). Similarly, the top most highly expressed miRNAs in normal-like/claudin-low cell lines were *hsa-miR-22, hsa-miR-532-3p*, *hsa-miR-125b*, *hsa-miR-501-5p*, and *hsa-miR-155** whereas in basal-like cell lines miRNAs of the miR-200 family, *hsa-miR-492*, *hsa-miR-26b*, *hsa-miR-617*, and *hsa-miR-155* were highly expressed (fold change $\geq$ 2) (Additional file 7.1: Table S8). Notably, *hsa-miR-155** and *hsa-miR-155*, which derived from the same precursor showed an opposing expression pattern in basal-like compared to normal-like/claudin-low BC cell lines.

Differential expression of miRNAs and *E cadherin* loss in breast cancer cell lines

E-cadherin elicits a growth suppressive effect in mammary epithelial cells and its expression is frequently lost in human breast tumors[29]. Loss of E-cadherin expression occurs either by gene mutation or by gene promoter hypermethylation. In this cohort of cell lines and in clinical specimens *E-cadherin* is preferentially mutated in luminal/ER-positive cell lines while it is frequently

inactivated by promoter hypermethylation in basal/ER-negative cell lines[21]. To determine whether loss of *E-cadherin* expression in human BC cell lines is associated with differential expression of miRNAs, we compared 10 *E-cadherin* mutant and 17 wild-type cell lines within the luminal-group of cell lines (Figure 7.4A) and 9 *E-cadherin* promoter hypermethylated with 8 non-methylated cell lines within the ER-negative/basal-group (Figure 7.4B). Our comparison revealed signatures of 24 and 31 differentially expressed miRNAs associated with *E-cadherin* mutant and wild type cell lines and with cell lines having *E-cadherin* promoter hypermethylation or not, respectively ($P < 0.05$) (Figure 7.4A and B; Additional file 7.1: Tables S9 and S10). Interestingly, both E-cadherin signatures had no miRNAs in common showing similar behavior.

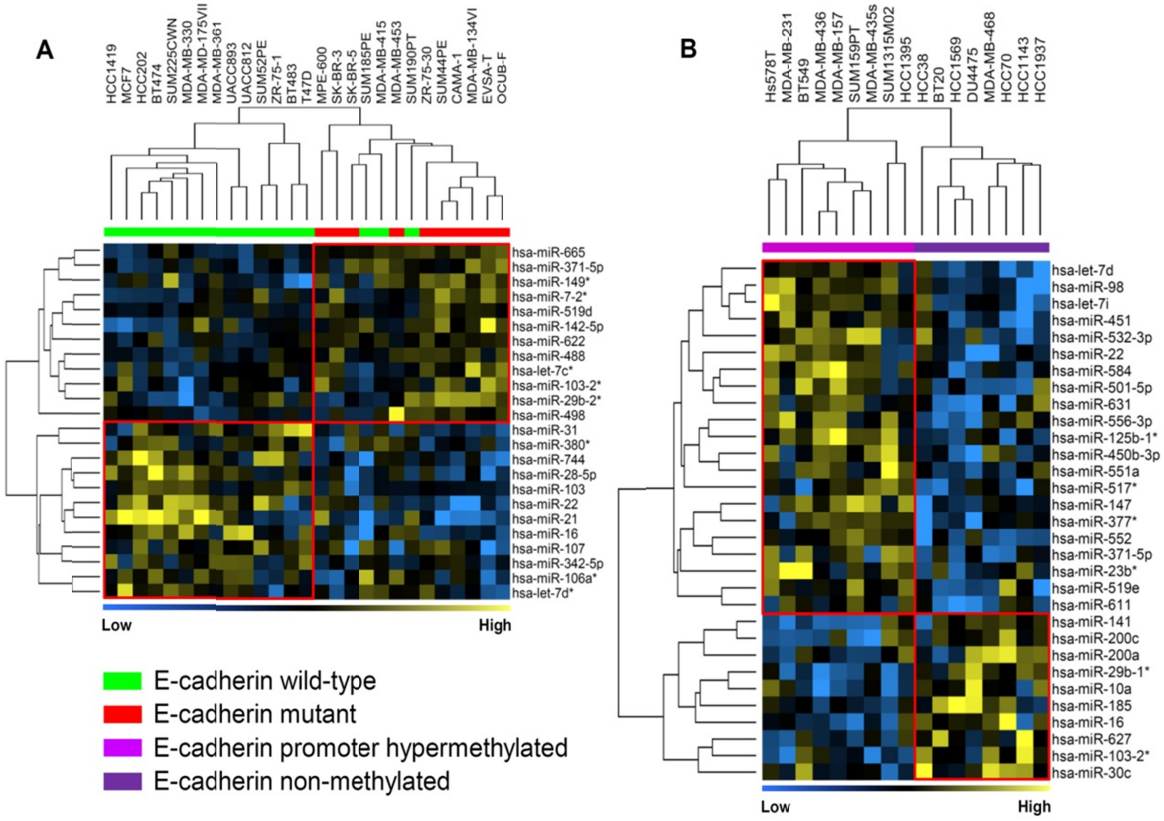


Figure 7.4 Differential expression of miRNAs with respect to *E-cadherin* status. **A.** Differentially expressed miRNAs between *E-cadherin* mutant and wild-type BC cell lines in luminal group and **B.** between *E-cadherin* promoter hypermethylated and wild-type BC cell lines in ER-negative/basal group. Cell lines HCC1806, SK-BR-7, SUM102PT, SUM149PT, and SUM229PE, which show partial promoter hypermethylation, were not included in the analysis since we were not sure to which extent partial promoter methylation will affect *E-cadherin* expression levels in these cell lines. The yellow and blue color indicates respectively high and low overall similarity of samples in miRNA expression.

Association of miRNAs expression with common genetic aberrations (*p16^{INK4a}*, *TP53*, *BRCA1*, *PIK3CA*, and *PTEN*)

The *BRCA1*, *p16^{INK4a}*, *TP53*, and *PTEN* genes play a growth suppressive role in mammary epithelial cells and are frequently inactivated in human breast tumor tissue. *PIK3CA* is also frequently mutated in breast cancer but has an oncogenic role. To reveal differentially expressed miRNAs associated with these mutations, we performed supervised analyses between mutant and wild-

type cell lines for these genes. To avoid the confounding effects of ER, we performed these analyses only in ER-positive or ER-negative cell lines. In our cohort, 13 cell lines had lost the *p16*^{INK4a} locus of which the majority are ER-negative. The comparison revealed that 30 miRNAs were differentially expressed between 9 p16^{INK4a} mutant and 7 wild-type cell lines in the ER-negative/basal-group (Figure 7.5 and Additional file 7.1: Table S11).

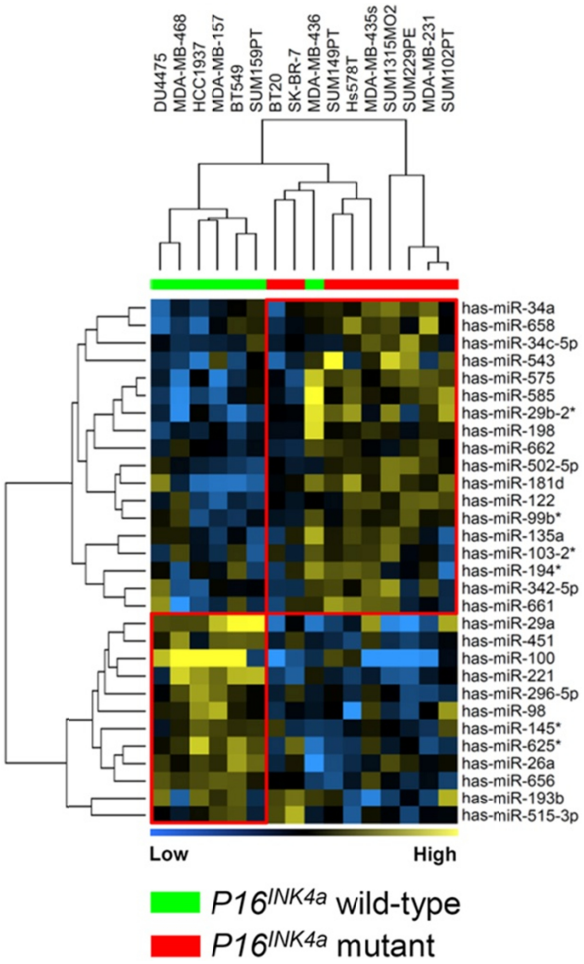


Figure 7.5 Differentially expressed miRNAs between *p16*^{INK4a} mutant and wild-type cell lines in the ER-negative/basal-group**.** The analysis was restricted to only ER-negative/basal-group of cell lines since majority of *p16*^{INK4a} mutant cell lines was ER-negative**.** The yellow and blue color indicates respectively high and low overall similarity of samples in miRNA expression.

The most highly expressed miRNAs in *p16*^{INK4a} mutant cell lines were *hsa-miR-29a* and *hsa-miR-100* (Fold change $\geq$ 2). Notably, the *p16*^{INK4a} miRNA signature in the ER-negative/basal-group of cell lines appeared to be pronounced and when used for clustering it could almost completely separate the cell lines mutant for the *p16*^{INK4a}gene from those that were wild-type. It should, however, be emphasized that the tumor suppressor genes *p16*^{INK4a} and *p14*^{ARF} are co-localized and are often simultaneously lost amongst cell lines. Therefore we cannot determine which of these genes might drive the differential expression of these miRNAs. Furthermore, the current cohort of BC cell lines contained 15 *PIK3CA* mutant (4 ER-negative, 11 ER-positive) and 8 *PTEN* mutant (4 ER-negative, 4 ER-positive) cell lines. Both *PIK3CA* and *PTEN* are components of the PI3K signaling pathway and mutated mutually exclusively in clinical specimens[30] and also in this cohort of cell lines[31]. Separate

analyses between *PIK3CA* or *PTEN* mutant versus wild-type cell lines for these genes did not yield a considerable number of differentially expressed miRNAs. We therefore combined all cell lines harboring mutations either in *PIK3CA* or *PTEN* and compared them with wild-type cell lines. As *PIK3CA/PTEN* mutations are present in both luminal- and ER-negative/basal-groups of the cell lines we therefore performed multiple analyses. First, we analyzed all cell lines and found 49 miRNAs, which were differentially expressed between mutant and wild-type cell lines (Additional file 7.2: Figure S2A and Additional file 7.1: Table S12). With successive analyses, we found 8 and 42 differentially expressed miRNAs among the ER-negative/basal-group and the luminal group, respectively (Additional file2: Figure S2A and Additional file1: Tables S13 and S14). With respect to *BRCA1*, which is mutated in 4 of the ER-negative cell lines, we identified 13 miRNAs being differentially expressed between 4 mutant and 12 wild-type BC cell lines (Additional file 7.2: Figure S2B and Additional file 7.1: Table S15). The two most highly expressed miRNAs in *BRCA1* mutant cell lines were *hsa-miR-29b* and *hsa-miR-891b* (Fold change $\geq$ 2). Finally, we analyzed differentially expressed miRNAs between *TP53* mutant (n = 39) and wild type cell lines (n = 9) and found 18 to be associated with the *TP53* mutation status (Additional file 7.2: Figure S3 and Additional file 7.1: Table S16). It is important to note however, that the differentially expressed miRNA signatures associated with *BRCA1*, *PIK3CA/PTEN*, and *TP53* mutations did not strongly discriminate between mutant and wild-type cell lines and therefore their significance may be of limited value.

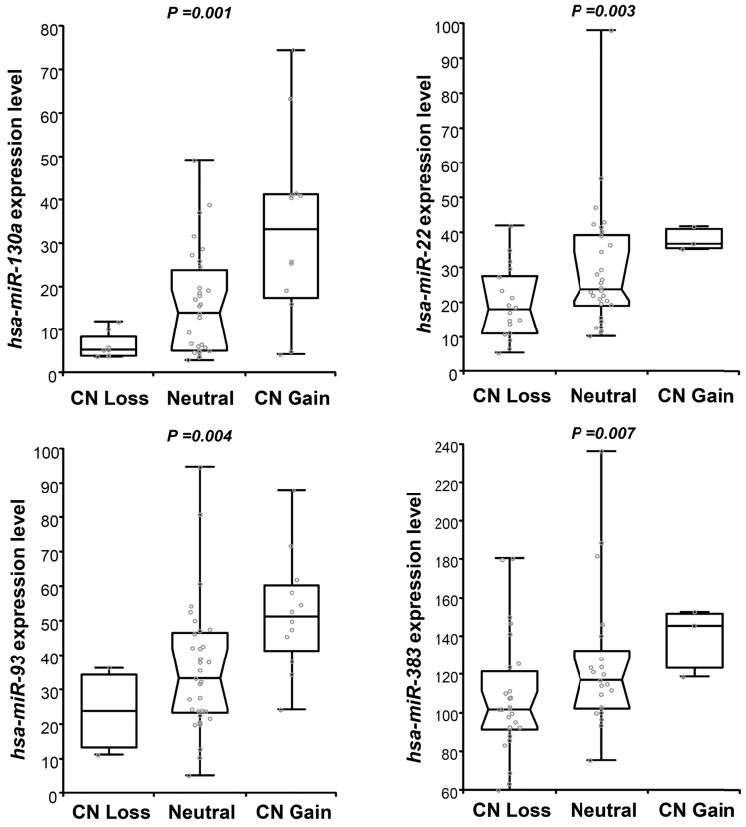


Figure 7.6 Association of miRNA expression with genomic copy number variation in BC cell lines. The top 4 most significant miRNAs are represented (see Additional file 7.1: Table S17 for a complete list). Kruskal-Wallis test was used to reveal significant associations of miRNAs with genome copy number variation. Expression levels of miRNA were presented on y-axis and the number of samples with copy number loss/gain or neutral is presented on x-axis. CN, copy number.

MiRNA genes show genomic aberrations in breast cancer cell lines

It is evident from previous studies that DNA copy number variations (CNVs) in breast tumor tissues can lead to differential expression of genes and miRNAs[32-34]. To investigate this in our cohort of BC cell lines, we first determined the DNA copy number variations of the cell lines by performing whole genome SNP profiling. These CNVs were then correlated with the expression levels of the 87 most variably expressed miRNAs in the cell lines. The correlation revealed 12 miRNAs, which were significantly associated with DNA CNVs (Kruskal-Wallis test, $P < 0.05$) (Additional file 7.1: Table S17). The top 4 most significantly associated miRNAs; *hsa-miR-130a* (11q12.1), *hsa-miR-22* (17p13.1), *hsa-miR-93* (7q22.1) and *hsa-miR-383* (8p22) with DNA CNVs in BC cell lines are shown in Figure 7.6.

DISCUSSION

Human BC cell lines are renewable resources that are extensively utilized as reliable workhorses to explore biological functions of clinically relevant molecules in breast cancer. An extensive molecular characterization and gene mutation analysis by us and others have suggested that BC cell lines have retained significant molecular features that are commonly observed in clinical breast tumors[35,36]. This prompted us to use 51 human BC cell lines as a discovery cohort to identify differentially expressed miRNAs between known intrinsic subtypes of breast cancer as well as those associated with common genetic aberrations present in the cell lines.

In this study, we demonstrated that global miRNA expression profiling can assign our cell line collection into two clusters. These two clusters predominantly mirror the ER-based dichotomy present in human BC cell lines, which may point to the fact that like mRNA, miRNA expression profiling allows the discrimination between ER-positive and ER-negative cell lines. This suggests that a significant number of miRNAs may be under the control of ER regulation. Interestingly, not all miRNA discriminating ER-positive and ER-negative cell lines are associated with ER status and therefore the division of the cell lines into two clusters may not be purely dictated by the expression of ER related miRNAs, but is likely that these miRNAs are also related to luminal versus basal cellular differentiation.

Our supervised analysis revealed a signature of 79 differentially expressed miRNAs between ER-positive and ER-negative cell lines, which included some of the previously known ER related miRNAs, such as a cluster of *hsa-miR-221/222*[12,37]. These miRNAs have been shown to interact directly with ER and cause a phenotypic shift from ER-positive to ER-negative tumor cells[37]. Our miRNA analyses in this cohort of cell lines therefore not only confirm some of the previous findings but also reveal new miRNAs (mentioned below), which may have potentially interesting biological roles in ER-driven cancer.

Among these potentially ER-related miRNAs, four miRNAs; *hsa-miR-26a, hsa-miR-92b, hsa-miR-191*, and *hsa-miR-492* appear to show consistently higher expression across all the ER-positive cell lines (Fold change $\geq$ 1.5) and as yet have not been implicated in breast carcinogenesis. Among them, *hsa-miR-26a* and *hsa-miR-92b* have previously been implicated in brain tumors[38,39]. The *hsa-miR-26* miRNA interacts with CDK4 and CENTG1 oncogenes and forms an integrated oncomir/oncogene

DNA cluster, which promotes glioblastoma tumor growth via RB1, PI3K/AKT, and JNK pathways[38]. On the other hand, *hsa-miR-92b* has been found to be exclusively over expressed in primary brain tumors but serves as a biomarker to discriminate brain primary cancer from metastasis[39]. The other two miRNAs, *hsa-miR-191* and *hsa-miR-492* have been linked to hepatic cancer. The increased expression of *hsa-miR-191* stimulates proliferation in hepatocellular carcinoma cell lines and its therapeutic targeting suppresses tumor masses *in vivo*[40]. The *hsa-miR-492* miRNA has been shown to be processed from keratin 19 gene and upregulated in metastatic hepatoblastoma[41]. How these miRNAs play their biological roles in the context of breast cancer remains unknown and demands clinical and functional validation studies.

Another major contribution to the overall miRNA profiling of this cell line cohort is the identification of a signature of 42 differentially expressed miRNAs, which discriminates between basal-like and normal-like/claudin-low BC cell lines. These miRNAs together with the ER-associated miRNAs are the major determinants of the overall clustering of the cell lines. Importantly, this signature includes all four members of *hsa-miR-200* family (*hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, hsa-miR-141*), *hsa-miR-155*, and *hsa-miR-622* miRNAs. Several studies have implicated these miRNAs to be involved in EMT, to be related to the stem cell like phenotype and to be associated with a switch in paclitaxel responsiveness[42-45]. In breast cancer, these miRNAs are known to regulate the EMT process by targeting the ZEB family of the transcription factors through an active negative feedback loop[46,47]. Importantly, the ZEB family of transcription factors was reported to be a repressor of E-cadherin expression in several epithelial carcinomas including breast carcinoma. Our finding that the *hsa-miR-200* cluster showed lower expression in normal-like/claudin-low BC cell lines fits with the fundamental literature that these miRNAs are indeed involved in EMT as the majority of the normal-like/claudin-low cell lines lack E-cadherin protein expression, exhibit low claudin expression, and generally display EMT like features as well as the BC stem cell phenotype ($CD44^{high}/CD24^{low}$)[48-50] and (Hollestelle *et al*, submitted). Other known significant miRNAs of this signature are *hsa-miR-155* and *hsa-miR-622*, which were also linked to enhanced tumorigenesis in various cancer types besides breast cancer[51,52]. In concordance with this, we recently found that the normal-like/claudin-low cell lines with low expression of these EMT-related miRNAs show highly aggressive growth characteristics *in vivo* when raised as xenografts in nude mice (Riaz *et al*, unpublished data). Importantly, this signature includes three previously unknown miRNAs (*hsa-miR-492*, *hsa-miR-26b*, *hsa-miR-617;* Fold change $\geq$ 1.5) to be associated with cell lines frequently showing EMT like characteristics.

We observed that the miRNA signatures associated with *E-cadherin* mutation and promoter hypermethylation include distinct miRNAs. This may point to an existence of unique biology at the miRNA level in tumors that show *E-cadherin* inactivation due to gene mutation rather than those that lose *E-cadherin* expression due to promoter hypermethylation. It is important to mention that *E-cadherin* promoter hypermethylated cell lines include all normal-like/claudin-low cell lines and most EMT-related miRNAs were also associated with *E-cadherin* promoter hypermethylation.

With respect to common genetic aberrations present in BC, our study reveals differentially expressed miRNA signatures associated with commonly mutated tumor suppressors and

oncogenes. Most significantly, a signature of 30 miRNAs associated with *p16^{INK4a}* mutant cell lines can strongly discriminate between mutant and wild-type ER-negative cell lines. *P16^{INK4a}* is a tumor suppressor gene, which is located on chromosomal band 9p21 that has been frequently altered in many human cancers [53]. It regulates the cell cycle by targeting CDK 4/6 through the pRb signaling pathway. Interestingly, this signature includes some miRNAs, which are involved in cell cycle regulation. For instance, miRNAs *hsa-miR-100*, which is highly expressed in the *p16^{INK4a}* wild-type cell lines, targets the *RBSP3* gene that in acute myeloid leukemia regulates the cell cycle through partial modulation of pRB/E2F1[54]. *Hsa-miR-34a* has been reported as a suppressor of cell proliferation and migration in colon cancer[55] and its mechanism of growth inhibition also involves cell cycle arrest followed by apoptosis[56]. Besides this, we found a few miRNAs to be only associated with mutation in *BRCA1, TP53* and *PIK3CA/PTEN*; these observed associations require further independent validation in breast cancer specimen.

Our finding that the differential expression of miRNAs is associated with ERBB2 overexpressing luminal-type cell lines is also intriguing since the miRNAs showing elevated expression in ERBB2-positive cell lines (Fold change $\geq$ 1.5) are not located on the ERBB2 amplicon (17q12) and thus may be regulated indirectly by genes in the ERBB2 amplicon. It is also important to mention that a majority of these miRNAs have already been implicated in various types of epithelial carcinoma including breast cancer[57-61]. It should however be noted that our study is associative and does not necessarily reveal causality. We therefore propose functional studies on these miRNAs to reveal more biological insights into their role regarding ERBB2 overexpression in breast cancer.

Finally, we also identified 12 miRNAs to be associated with CNVs in BC cell lines (Additional file 7.1: Table S17). Majority of these miRNAs (*hsa-miR-130a*, *hsa-miR-93*, *hsa-miR-383*, *hsa-miR29c*, *hsa-miR-382*, *hsa-miR-31*) were already found to be located in regions that exhibited DNA copy number abnormalities in breast cancer tumors[62]. Importantly, this repertoire of miRNAs includes *hsa-miR-22* previously shown to be regulated by ER[63] and we provide evidence that it can also be regulated by the loss of the locus containing this miRNA. Six miRNAs of this repertoire are located in the genomic regions containing known protein coding. All miRNAs of this repertoire have also been implicated in various cancers but a thorough validation of these miRNAs with respect to DNA copy number changes in clinical specimen is imperative.

CONCLUSIONS

In summary, our analyses show that the dichotomy in breast cancer cell lines is, in line with the general consensus in the field, related to ER status. However, part of this dichotomy may be related to ER-independent luminal versus basal differentiation status of cell lines. Secondly, another prominent dichotomy observed among BC cell lines discriminating basal-like from normal-like/claudin-low cell lines involves miRNAs related to EMT and/or stemness. Furthermore, we reveal sets of miRNAs associated with genes frequently amplified (ERBB2) or mutated (*p16^{INK4a}*, *PIK3CA* and/or *PTEN, E-cadherin*, *BRCA1)* in BC cell lines and thus these miRNAs may contribute to the function of these oncogenes/tumor suppressors. Finally, certain miRNAs themselves are located in

genomic regions, which frequently show genetic aberrations implying that these miRNAs themselves may have a potential driver role or contribute to known drivers in these genomic regions. Our current findings call for further validation of these signatures in clinical specimens. Importantly, our study provides a unique molecular miRNA portrait of human breast cancer cell lines, which can be exploited for functional studies of clinically important miRNAs.

ABBREVIATIONS

ATCC, american type culture collection; BC, breast cancer; BRB, biometric research branch; BRCA1, breast cancer 1; BST2, bone marrow stromal cell antigen 2; CK, cytokeratin; CDKs, cyclin-dependent kinases; CNVs, copy number variations; DSMZ, deutsche sammlung von mikroorganismen und zellkulturen; EMT, epithelial to mesenchymal transition; ER, estrogen receptor-α; ERBB2, erythroblastic leukemia viral oncogene homolog 2; ESR1, estrogen receptor; EGFR, epidermal growth factor receptor; FABP7, fatty acid binding protein 7; FOXA1, forkhead box protein A1; GEO, gene expression omnibus; GRB7, growth factor receptor-bound protein 7; JNKs, c-Jun N-terminal kinases; KRT, keratin; PGR, progesterone receptor; PI3K, phosphoinositide 3-kinase; PTEN, phosphatase and tensin homolog; RB, retinoblastoma protein; STR, short tandem repeat; STARTD3, StAR-related lipid transfer domain protein 3; TFF3, trefoil factor 3; TP53, tumor protein 53; UTR, untranslated region.

COMPETING INTERESTS

All authors declared that they have no competing interest.

AUTHORS' CONTRIBUTIONS

JAF, JWMM, EAW, and MR designed the study. MR, JAF and JWMM wrote the manuscript. MTMV and AWMB performed the miRNA expression analysis. JL and AH did gene mutation analysis. JH performed STR analysis. BO and WJCP did RNA isolation and quantification. AAAJH, MS, and MR performed the statistical data analyses. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

Janssen Pharmaceutical, a Johnson and Johnson Company, Beerse, Belgium, sponsored a part of the molecular profiling performed within this study. This study was supported financially by the Netherlands Genomics Initiative (NGI) / Netherlands Organization for Scientific Research (NWO) and by the Dutch Cancer Society (EMCR 2007-3794).

ADDITIONAL FILES

Additional file 7.1

Excel file of Tables S1-S17 containing lists of differentially expressed miRNAs associated with breast cancer molecular subtypes and recurrent genetic aberrations.

Additional file 7.2

Figure S1, S2, S3

Additional files are available on:

https://www.dropbox.com/sh/4fwrk0lit91dsp5/WP1sZYZAQw?m

REFERENCES

1. Pasquinelli, A.E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* **13**, 271-82 (2012).
2. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
3. Alvarez-Garcia, I. & Miska, E.A. MicroRNA functions in animal development and human disease. *Development* **132**, 4653-62 (2005).
4. Zhao, Y. & Srivastava, D. A developmental view of microRNA function. *Trends Biochem Sci* **32**, 189-97 (2007).
5. Enerly, E. et al. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* **6**, e16915 (2011).
6. Hwang, H.W. & Mendell, J.T. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *Br J Cancer* **94**, 776-80 (2006).
7. Leung, A.K. & Sharp, P.A. MicroRNA functions in stress responses. *Mol Cell* **40**, 205-15 (2010).
8. Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**, 586-91 (2009).
9. Blenkiron, C. et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **8**, R214 (2007).
10. Iorio, M.V. et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65**, 7065-70 (2005).
11. Van der Auwera, I. et al. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* **103**, 532-41 (2010).
12. Foekens, J.A. et al. Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proc Natl Acad Sci U S A* **105**, 13021-6 (2008).
13. Camps, C. et al. hsa-miR-210 Is induced by hypoxia and is an independent prognostic factor in breast cancer. *Clin Cancer Res* **14**, 1340-8 (2008).

14. Tavazoie, S.F. et al. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147-52 (2008).
15. Rodriguez-Gonzalez, F.G. et al. MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. *Breast Cancer Res Treat* **127**, 43-51 (2011).
16. Jansen, M.P. et al. High miR-26a and low CDC2 levels associate with decreased EZH2 expression and with favorable outcome on tamoxifen in metastatic breast cancer. *Breast Cancer Res Treat* (2011).
17. Ryu, S. et al. Discovery of novel human breast cancer microRNAs from deep sequencing data by analysis of pri-microRNA secondary structures. *PLoS One* **6**, e16403 (2011).
18. Farazi, T.A., Spitzer, J.I., Morozov, P. & Tuschl, T. miRNAs in human cancer. *J Pathol* **223**, 102-15 (2011).
19. Miller, T.E. et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *J Biol Chem* **283**, 29897-903 (2008).
20. Voorhoeve, P.M. et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124**, 1169-81 (2006).
21. Hollestelle, A. et al. Four human breast cancer cell lines with biallelic inactivating alpha-catenin gene mutations. *Breast Cancer Res Treat* **122**, 125-33 (2010).
22. Sieuwerts, A.M. et al. How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin Cancer Res* **11**, 7311-21 (2005).
23. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747-52 (2000).
24. Pothof, J. et al. MicroRNA-mediated gene silencing modulates the UV-induced DNA-damage response. *EMBO J* **28**, 2090-9 (2009).
25. Hollestelle, A. et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* **121**, 53-64 (2010).
26. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-4 (2004).
27. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-74 (2001).
28. Zeng, Y.A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568-77 (2010).
29. Vos, C.B. et al. E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer* **76**, 1131-3 (1997).
30. Saal, L.H. et al. PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res* **65**, 2554-9 (2005).
31. Hollestelle, A., Elstrodt, F., Nagel, J.H., Kallemeijn, W.W. & Schutte, M. Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol Cancer Res* **5**, 195-201 (2007).

32. Pollack, J.R. et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* **99**, 12963-8 (2002).
33. Hyman, E. et al. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* **62**, 6240-5 (2002).
34. Calin, G.A. et al. 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 (2004).
35. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-27 (2006).
36. Kao, J. et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4**, e6146 (2009).
37. Di Leva, G. et al. MicroRNA cluster 221-222 and estrogen receptor alpha interactions in breast cancer. *J Natl Cancer Inst* **102**, 706-21 (2010).
38. Kim, H. et al. Integrative genome analysis reveals an oncomir/oncogene cluster regulating glioblastoma survivorship. *Proc Natl Acad Sci U S A* **107**, 2183-8 (2010).
39. Nass, D. et al. MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. *Brain Pathol* **19**, 375-83 (2009).
40. Elyakim, E. et al. hsa-miR-191 is a candidate oncogene target for hepatocellular carcinoma therapy. *Cancer Res* **70**, 8077-87 (2010).
41. von Frowein, J. et al. MicroRNA-492 is processed from the keratin 19 gene and up-regulated in metastatic hepatoblastoma. *Hepatology* **53**, 833-42 (2011).
42. Gregory, P.A. et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593-601 (2008).
43. Korpal, M., Lee, E.S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**, 14910-4 (2008).
44. Adam, L. et al. miR-200 expression regulates epithelial-to-mesenchymal transition in bladder cancer cells and reverses resistance to epidermal growth factor receptor therapy. *Clin Cancer Res* **15**, 5060-72 (2009).
45. Gupta, P.B. et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**, 645-59 (2009).
46. Hurteau, G.J., Carlson, J.A., Spivack, S.D. & Brock, G.J. Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. *Cancer Res* **67**, 7972-6 (2007).
47. Christoffersen, N.R., Silahtaroglu, A., Orom, U.A., Kauppinen, S. & Lund, A.H. miR-200b mediates post-transcriptional repression of ZFHX1B. *RNA* **13**, 1172-8 (2007).
48. Park, S.M., Gaur, A.B., Lengyel, E. & Peter, M.E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* **22**, 894-907 (2008).

49. Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* **12**, R68 (2010).
50. Sarrio, D., Franklin, C.K., Mackay, A., Reis-Filho, J.S. & Isacke, C.M. Epithelial and mesenchymal subpopulations within normal basal breast cell lines exhibit distinct stem cell/progenitor properties. *Stem Cells* **30**, 292-303 (2012).
51. Guo, X.B. et al. Down-regulation of miR-622 in gastric cancer promotes cellular invasion and tumor metastasis by targeting ING1 gene. *World J Gastroenterol* **17**, 1895-902 (2011).
52. Donnem, T. et al. Prognostic impact of MiR-155 in non-small cell lung cancer evaluated by in situ hybridization. *J Transl Med* **9**, 6 (2011).
53. Sherr, C.J. Cancer cell cycles. *Science* **274**, 1672-7 (1996).
54. Zheng, Y.S. et al. MiR-100 regulates cell differentiation and survival by targeting RBSP3, a phosphatase-like tumor suppressor in acute myeloid leukemia. *Oncogene* **31**, 80-92 (2012).
55. Tazawa, H., Tsuchiya, N., Izumiya, M. & Nakagama, H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci U S A* **104**, 15472-7 (2007).
56. Cole, K.A. et al. A functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene. *Mol Cancer Res* **6**, 735-42 (2008).
57. Redova, M. et al. Circulating miR-378 and miR-451 in serum are potential biomarkers for renal cell carcinoma. *J Transl Med* **10**, 55 (2012).
58. Zhao, J.J. et al. microRNA expression profile and identification of miR-29 as a prognostic marker and pathogenetic factor by targeting CDK6 in mantle cell lymphoma. *Blood* **115**, 2630-9 (2010).
59. Kriegel, A.J., Liu, Y., Fang, Y., Ding, X. & Liang, M. The miR-29 family: genomics, cell biology, and relevance to renal and cardiovascular injury. *Physiol Genomics* **44**, 237-44 (2012).
60. Jedlinski, D.J., Gabrovska, P.N., Weinstein, S.R., Smith, R.A. & Griffiths, L.R. Single nucleotide polymorphism in hsa-mir-196a-2 and breast cancer risk: a case control study. *Twin Res Hum Genet* **14**, 417-21 (2011).
61. Vrba, L. et al. Role for DNA methylation in the regulation of miR-200c and miR-141 expression in normal and cancer cells. *PLoS One* **5**, e8697 (2010).
62. Zhang, L. et al. microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A* **103**, 9136-41 (2006).
63. Pandey, D.P. & Picard, D. miR-22 inhibits estrogen signaling by directly targeting the estrogen receptor alpha mRNA. *Mol Cell Biol* **29**, 3783-90 (2009).

Chapter 8

Growth and metastatic behavior of human breast cancer cell lines in mice

Muhammad Riaz, Buddy Setyono-Han†, Mieke A. Timmermans, Anita M. Trapman, Joan Bolt-de Vries, Antoinette Hollestelle, Maxime P. Look, John A. Foekens, and John W.M. Martens.

In preparation

ABSTRACT

Introduction Conventional and molecular profiling has shown that breast cancer (BC) is a group of multiple diseases with a considerable amount of intra- and inter-tumor heterogeneity. This heterogeneity has led to the subdivision of BC into several molecular subtypes, each with a distinct type of response to therapy, metastatic potential and prognosis. Human BC cell lines are available as model systems; however few are used as *in vivo* experimental models. Thus, models representing the wide variety of breast cancer phenotypes in the clinic are largely not available. Here, we explored the *in vivo* growth potential and metastatic behavior of 24 human BC cell lines representing four molecular subtypes to generate useful models for experimentation and to determine whether these cell lines recapitulate in mice their subtype-related biological behavior.

Methods To study and image biological behavior of BC cell lines in mice, the cells were transduced using lentivirus particles with a luciferase reporter gene. Subsequently, per cell line 3 independent monoclonal clones with similar luciferase activity and molecular features comparable to the parental lines were generated and implanted orthotopically under the mammary fat pad of nude mice. Xenograft tumor growth was monitored by measuring tumor size twice a week for a period of 12 months. At the end of the experiment, using the IVIS Imaging System, tumor size and presence of distant organ metastases were assessed by measuring the tumor bioluminescence. Subtype specific differences in tumor growth and mice survival were compared using a Kruskal-Wallis test and Cox regression analysis. A P-value < 0.05 was considered to be statistically significant.

Results 87.5% of the cell lines gave rise to xenografts in the mammary fat pad of recipient mice. A clear difference in BC xenograft growth and tumor-related survival was observed between molecular subtypes (P =0.01 Kruskal-Wallis test). Normal-like/claudin-low and basal-like BC cell lines were found to be more aggressive in mice and showed poor tumor-related survival (HR: 6.89; 95% CI: 1.96-24.22; P =0.003 and HR: 4.30; 95% CI: 1.11-16.75; P =0.04, respectively). The most common sites colonized by BC cell lines in mice were lymph-nodes, lungs, liver, and ovaries. Relapses to other sites such as bone, brain, and pancreas were rarely observed during the maximum of 12 months of the experiments. No clear association between BC subtype and type of distant organ metastasis was observed.

Conclusions We established a panel of 21 xenografts from a molecularly diverse collection of human BC cell lines. Comparable to clinical behavior in human, we observed that normal-like/claudin-low cell lines were more aggressive in mice than the cell lines of other molecular subtypes. Although no relationships between subtype and site of relapse was observed, a large variety of clinical manifestations of clinical breast cancers is represented in this panel of BC cell line xenografts, which indicates that this panel can serve as very useful preclinical model for studying the biology of these phenotypes and for screening for interfering drugs.

INTRODUCTION

Breast cancer is not a single disease but rather a group of multiple distinct diseases with an extensive intra- and inter-tumor heterogeneity. The heterogeneous nature of breast tumors has been captured histologically decades ago by pathologists and by clinicians during clinical examinations. Based on such differences breast tumors were classified into various histological subtypes with a different prognosis [1]. The classical pathology-based classification has been refined and at times replaced by modern molecular classifications, which have the potential to combine disease mechanisms with clinical outcome measures [2]. Molecular classification, an ongoing process, is currently mainly driven by differences in the expression of several thousands of genes detectable in the human breast tissues, and such changes are generally measured using gene expression microarray technology[3].

Using cDNA microarrays, five main molecular subtypes have initially been identified[3]. These subtypes include two luminal-like subtypes: luminal A and luminal B subtypes, an ERBB2 overexpressing subtype, a basal-like subtype, and a less well defined normal-like subtype. The existence of these five subtypes in breast cancer is very stable and has been observed and validated by a number of subsequent studies in independent series of tumors from breast cancer patients with different ethnic background[4]. These subtypes not only display distinct gene expression patterns, they also show a significant difference in disease-free as well as overall survival[5]. Additionally, these subtypes show distinct preferences for organ-specific metastases[6,7]. For instance, the luminal subtypes of breast cancer preferentially metastasize to bone and the basal-like subtype more frequently metastasize to lungs and brain[6]. Furthermore, the stability of the genome as well as the response to certain types of chemotherapy of these subtypes is also diverse[8].

In the past decade, the molecular classification of breast cancer has been further refined and based on similarity in specific gene expression patterns more molecular subtypes have been identified[9,10]. Recently, Lehmann and colleagues identified 6 molecular subtypes within triple-negative breast cancer (ER-, PGR-, ERBB2-)[11]. This grouping of breast tumors based on gene expression analysis allowed the identification of several subtype-specific molecular drivers, which have been successfully targeted therapeutically using human BC cell lines as model systems[11]. Human BC cell lines are well-known and widely used model systems to understand breast cancer biology and pharmacology. Recently, a large collection of cell lines including BC cell lines have been used as *in vitro* pre-clinical model systems to uncover new biomarkers of sensitivity to the currently applied therapeutic drugs[12,13]. Furthermore, accumulating evidence suggests that the currently available BC cell lines mirror the molecular portraits of primary breast tumors. For instance, BC cell lines cluster into luminal-type, ERBB2 positive, basal-like, and normal-like/claudin low molecular subtypes, as do primary tumors[14]. Additionally, the cell lines exhibit a comparable degree of heterogeneity in copy number and expression abnormalities as observed in primary tumors, and they carry many of the genetic abnormalities commonly found in primary breast tumors[15].

The interpretation and extrapolation of cell-line related data, however, require an understanding of how uniformly and stably these cell lines mimic primary breast tumors at both *in vitro* and *in vivo* levels. Previous studies have suggested that the cell lines, despite of considerable similarities, also show some discrepancies with clinical breast tumors. These discrepancies include higher frequency of *TP53* gene mutation and over representation of ER-negative cell lines compared with human clinical breast tumors. In this study we explored in nude mice the *in vivo* growth potential and metastatic capacity of a panel of 24 human breast cancer cell lines representing four major molecular subtypes.

MATERIALS AND METHODS

Cell lines culture conditions

The 24 human breast cancer cell lines used in this study are listed in Table 8.1. Cells were maintained in RPMI-1640 medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) and antibiotic agents (100 µg/mL penicillin G and 80 µg/mL streptomycin (P/S)) in a humidified atmosphere of 5% CO2 and 95% air at 37°C. For experiments, cells were seeded in T-162 tissue culture flasks in normal culture medium containing 10% FBS until the cultures were 70-80% confluent.

Table 8.1 Molecular characteristics of human breast cancer cell lines

Cell lines	Protein expression				Molecular subtype
	ER	PGR	ERBB2	EGFR	
CAMA-1	+	+	–	–	Luminal
MCF-7	+	+	–	–	Luminal
MDA-MB-175VII	+	–	–	–	Luminal
SUM52PE	+	–	–	–	Luminal
T47D	+	+	–	+	Luminal
ZR75-1	+	+	–	–	Luminal
EVSA-T	–	+	Amp	–	ERBB2-positive
MDA-MB-453	–	–	Amp	–	ERBB2-positive
SK-BR-3	–	+	Amp	–	ERBB2-positive
UACC812	+	+	Amp	–	ERBB2-positive
UACC893	–	–	Amp	+	ERBB2-positive
BT20	–	–	–	+	Basal-like
HCC1937	–	–	–	+	Basal-like
MDA-MD-468	–	–	–	+	Basal-like
SUM229PE	–	–	–	+	Basal-like
BT549	–	–	–	+	Normal-like/Claudin-low
Hs578T	–	–	–	+	Normal-like/Claudin-low
MDA-MB-231	–	–	–	–	Normal-like/Claudin-low
MDA-MD-436	–	–	–	+	Normal-like/Claudin-low
SK-BR-7	–	–	–	+	Normal-like/Claudin-low
SUM159PT	–	–	–	+	Normal-like/Claudin-low
SUM149PT	–	–	–	+	Normal-like/Claudin-low
SUM1315MO2	–	–	–	+	Normal-like/Claudin-low
DU4475	–	–	–	+	Other

+, protein expression; –, no expression; Amp, amplification

Lentivirus production

Lentivirus-containing luciferase construct was produced as described below. On day 1, 8-10 $\times 10^6$ HEK293T cells were seeded in 100 mm $\times$ 20 mm tissue culture petri dishes using 14 mL RPMI-P/S culture medium. On day 2, cells were transfected using polyethyleneimine (PEI) transfection reagent. For each petri dish, a DNA mixture of viral plasmid and three helper plasmids were prepared in 0.5 mL of HBS (150 mM NaCl, 20 mM HEPES, pH 7.3) in a 50 mL falcon tube. The DNA mixture contained 20 µg of the viral vector with the luciferase gene (Cag-Luc), 10 µg of pMDLg/RRE, 5 µg of pRSV-REV [16], and 5 µg of pVSVG [17]. Then, PEI solution was prepared by adding 120 µg of PEI in 0.5 mL of HBS. The transfection mixture was then prepared by mixing PEI solution dropwise to the DNA mixture (1:4 µg; PEI and DNA) and subsequent incubation for 15 min at room temperature[18]. The medium in each dish was replaced with serum-free culture medium. Then, the transfection mixture was added dropwise to the HEK cells while swirling the dish to allow homogenous mixing. After 3 h of incubation, the medium-containing transfection mixture was replaced with 10 mL fresh culture medium and incubated for 24 h.

On day 3, after 24 h of transfection, the old medium was replaced with 6 mL fresh medium and incubated for an additional 24 h. On day 4, at 48 h, the medium-containing virus particles was collected from all dishes into a 50 mL tube and 6 mL fresh medium was added to the dishes. On day 5, at 72 h of transfection, the medium-containing viruses was again collected. The media collected at 48 h and 72 h were pooled, centrifuged at 300 $\times$ g for 5 min, and passed through a 0.45-µm PVDF filter to eliminate cellular debris. The virus particles were concentrated by adding PEG-it virus precipitation solution (System Bioscience) to the supernatant-containing virus particles with 1:4 by volume. The mixture was incubated for 24 h at 4°C and then centrifuged at 1,500 x g for 30 min at 4°C to precipitate lentivirus particles. Finally, the precipitated virus particles were resuspended in 200 µL of cold sterile Phosphate Buffered-Saline (PBS). Aliquots of 50 µL were stored at -80°C for future transduction experiments.

Transduction and clone selection

The cells to be transduced were seeded in a 24-well tissue culture plate in 0.5 mL culture medium. When the culture was approximately 50-60% confluent, 50 µL of concentrated virus particles were added to the well and incubated for 6-8 h at 37°C to allow transduction. After transduction, the transduced cells were trypsinized, subjected to the serial dilution from theoretically 50,000 cells to 2 cells per mL in normal culture medium and all dilutions were seeded in 96-well round-bottomed plates with 150 µL volume per well. To collect monoclonal populations of the cells, only plates in which colonies of cells were theoretically grown out from a single seeded cell were used. These colonies were screened for luciferase expression by adding D-luciferin (150 mg/mL medium) in the well and then bioluminescence was measured using the IVIS camera system (Xenogen, Alameda, CA, USA) as explained below. Multiple bioluminescent clones of each cell line were harvested and the clones showing similar luciferase activity with each other were further propagated. Additionally, a morphological and molecular comparison was performed between the propagated clones and their corresponding parental cell lines. The clones showing similar phenotypic and

molecular characteristics with each other and with the corresponding parental cell lines were selected for implantation experiments. The flow chart of the complete methodology is presented in Figure 8.1.

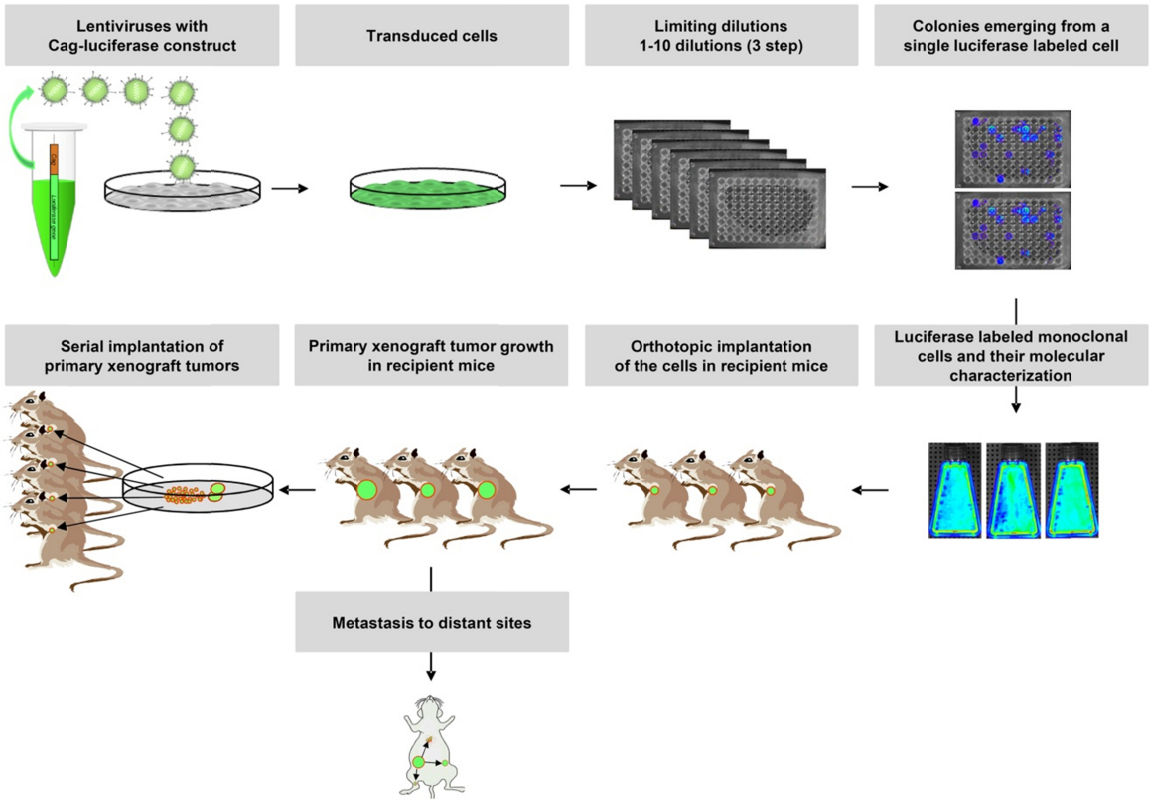


Figure 8.1 Schematic diagram showing methodology involving cell line labeling with the luciferase gene, clone selection and their characterization for orthotopic implantation in mice.

Cytospin staining and Immunohistochemistry

ER, PGR, ERBB2 and EGFR protein expression staining of the selected clones was performed on cytospins. Prior to cytospin, the clones were cultured in T75 tissue culture flasks and between 50-60% confluency the culture medium was aspirated, cells were washed 2 times with cold PBS and then scraped with a rubber policeman in an excessive volume of cold PBS. The scraped cells were then centrifuged and the cell pellet was resuspended in 25 ml 2% formalin solution for overnight fixation. The next day, the cells were centrifuged, the supernatant was removed and the cells were resuspended in PBS (1×10^6/mL). For the cytospin procedure, glass slides along with the paper pads and the cuvettes were mounted on the metallic cytospin holder. To make the paper pad wet, a pre-run was performed by loading 100 µL of PBS to the cuvettes. Then, 100 µL cell suspension was loaded and centrifuged at 130 g for 5 min. After centrifugation, the glass slides containing fresh cytospinned cells were separated from the cytospin holder and subjected to haematoxylin and eosin (H & E) staining as well as for ER, PGR, ERBB2, and EGFR protein staining.

For immunohistochemistry (IHC) experiments, primary xenograft-tumors and mice body organs with micrometastases were fixed in 10% buffered formalin solution for 24 h and then processed for paraffin embedding, sectioning, and H & E staining as well as ER, PGR, ERBB2 and EGFR protein staining. Primary antibody of ER (clone 1D5; M7047; Dako, Glostrup, Denmark) was diluted at 1:40,

PGR (A0098; Dako; Glostrup, Denmark) was diluted at 1:50, ERBB2 ready to use kit (HercepTest™; Dako, Glostrup, Denmark), and EGFR ready to use kit (EGFR pharmDx™; Dako, Glostrup, Denmark), were used for the cytospins. On the cytospins, these antibodies were developed using Envision™ Plus system (mouse, K4006; rabbit, K4008, Dako, Glostrup, Denmark). In the xenografts, the ER primary antibody was developed using animal research kit horse reddish peroxidise (K3954, Dako, Glostrup, Denmark). For EGFR expression in the xenograft, a Rabbit EGFR (D38B1; Cell Signaling, BZ Leiden, The Netherlands) diluted at 1:100 was used, and for PGR and ERBB2, rabbit primary antibodies at same dilution as described for cytospin staining were used.

Mouse strain and animal care

NMRI/Nu-Nu nude female mice aged 6-8 week (Charles River Laboratories, USA) were used in this study. Experiments were carried out according to the approved animal research protocol [EMC 2541 (136-11-03)] by the Institutional Animal Ethical Committee (DEC), Erasmus University Medical Center, Rotterdam, the Netherlands.

Cell line implantation and tumor size determination

Three mice were used to implant three selected clones from each cell line. The mice were anesthetized by the exposure of 2% isoflurane and placed ventral side up on a pre-warmed injecting pad. Approximately 10-12 million cells from each clone suspended in 200 µL fresh culture medium with 20% serum were implanted under the mammary fat pad of the 4th nipple (either left or right side of the body) with a 1-cc syringe and 24-gauge needle. Two weeks after implantation, xenograft tumor viability was checked by bioluminescence as explained below. The size of xenografts was measured twice a week using a vernier calliper [xeno-tumor size (mm^2) = length x width] and the mice were followed for a total period of maximal 12 months until they were sacrificed. To allow growth of ER-positive cell lines, a pellet of β-estradiol (Innovative Research of America) was implanted under the skin on the dorsal side of the neck of the mouse. Due to ethical reasons, primary xenograft tumors reaching the size of 225 mm^2 were surgically removed while leaving an approximately 16 mm^2 portion on the site to further facilitate invasion and metastasis. After the follow-up period, the intact animals alive were imaged to visualize distant organ metastasis. Subsequently, animals were sacrificed and body organs were excised. The excised body organs were re-imaged to assess organ micrometastases. The primary tumors and body organs were preserved in 10% formalin solution (Sigma, St, Missouri, USA) for subsequent histological analyses..

Bioluminescence imaging

Bioluminescence imaging was performed using IVIS Imaging System (Xenogen product from Caliper Life Sciences) according to manufacturer's instructions. Briefly, for *in vivo* imaging, at indicated time points, the substrate 200 µL of D-luciferin (10 µg/gram of mice weight) was injected into the mice body intraperitoneally approximately 10-15 min before imaging. The mice were then anesthetized with 2% isoflurane and placed onto the warmed stage inside the light-tight camera

box with a continuous exposure of 2% isoflurane. Bioluminescence images were acquired with an acquisition time ranging from 0.1 sec to 5 min depending upon the cell line used. For *ex vivo* imaging of cells, 150 µg/mL of D-luciferin in culture media was added to each well of a 96-well plate that contained cells, and the plate was imaged for 10-40 sec. For *ex vivo* imaging of mice tissues (body organs), 150 µg/gram D-luciferin was injected into the mouse body before necropsy. For imaging of body organs, 300 µg/mL of D-luciferin in PBS was used in a petri dish containing the excised organs. The imaging time ranged from 0.1 sec to 5 min. The body organs were subsequently fixed in 10% formalin and prepared for standard histopathology. Imaging and quantification of signals were controlled by acquisition and analysis software Living Imaging 3.0 (Xenogen), and data was expressed as total number of photon/sec/cm^2.

Data analyses

Cox regression analysis was used to reveal subtype-specific tumor-related survival in the mice. Kruskal-Wallis and Mann-Whitney tests were used to evaluate significant differences in growth rate of the cell lines across the intrinsic subtype. A P-value of < 0.05 was considered to be statistically significant.

RESULTS AND DISCUSSION

Implantation of 24 breast cancer cell lines in mice

To follow BC cells in the body of the mice, all cell lines were first stably transduced with a luciferase reporter gene using lentiviruses as vehicles. The luciferase expression was driven by the CAG promoter, which was previously shown to be a potent driver of the transgene expression in stem cell differentiation[19,20]. To overcome variable copy numbers of the luciferase gene due to random integration of the transgene by the lentiviruses, from each cell line 3 clones with equivalent luciferase activity were selected. Additionally, molecular characterizations of the selected clones were performed and it appeared that these clones showed a similar expression of ER, PGR, ERBB2, and EGFR proteins compared with their respective parental cell lines (Supplementary Figure 8.1). To determine the intrinsic molecular subtype-specific growth characteristics of the BC cell lines *in vivo*, 6 luminal-subtype, 5 ERBB2-positive, 4 basal-like, and 8 normal-like/claudin-low subtype cell lines were implanted in nude mice. In addition, 2 clones of the DU4475 cell line, which, according to microarray gene expression analysis did not cluster to either of the known intrinsic subtypes (chapter 7; Figure 7.1) and previously proposed to have cancer stem cell-like features due to activation of the Wnt signaling pathway[14], were implanted.

Subtype specific growth characteristics of BC xenografts

All 6 luminal-type cell lines, which included CAMA-1, MCF-7, MDA-MD-175VII, SUM52PE, T47D, and ZR-75-1 showed growth at the primary site in mice. However, for 5 of these 6 cell lines [CAMA-1 (2 of 3 clones), MDA-MD-175VII (2 of 3 clones), SUM52PE (2 of 2 clones), T47D (3 clones), and ZR-75-1 (3 clones)], a dramatic loss in body weight within one month after cell implantation was observed.

Therefore, these mice had to be taken off study and were censored at this time point in survival analyses. Interestingly, one clone of the CAMA-1 cell line at month 10 and all three clones of the MCF-7 cell line at month 1, 5 and 8, respectively, successfully gave rise to primary tumor xenografts of the maximal allowed size of 225 cm^2 (Supplementary Figure 8.2; Luminal-like cell lines). The cause of loss of body weight in mice bearing luminal-type xenografts remains unknown as pathological and microbiological analyses of these mice did not reveal any peculiar cell line-related concerns. It is however important to note that all the luminal-type cell lines are ER-positive [14] and therefore required a suitable concentration of estrogen for their steady growth under these *in vivo* conditions. To provide sufficient estrogen for optimal growth of these cell lines in mice, an estrogen pellet was implanted subcutaneously to each mouse 2 days before implantation of the cells. As this weight loss phenomenon was not observed in mice not receiving an estrogen pellet, a surplus of released estrogen above the physiological level in the mice carrying the luminal xenografts could have caused the unexpected decline in body weight.

Among the 5 ERBB2-positive cell lines (EVSA-T, MDA-MB-453, SK-BR-3, UACC812, and UACC893), all except MDA-MB-453 showed primary tumor growth in mice albeit with variable growth rates. As a whole, UACC812 and UACC893 cell lines exhibited the fastest growth rate compared with the other 3 ERBB2-positive cell lines (Supplementary Figure 8.2; ERBB2-positive cell lines). Similarly, xenograft tumors from all 4 basal-like cell lines (BT20, HCC1937, MDA-MB-468, and SUM229PE) were successfully established. Notably, only 1 out of the 3 BT20 implanted clones could grow in the mouse mammary fat pad while the other two regressed after establishing xenograft tumors of 36 mm^2 within two months post-implantation (Supplementary Figure 8.2, Basal-like cell lines).

Among 8 normal-like/claudin-low cell lines (BT549, Hs578T, MDA-MB-231, MDA-MB-436, SK-BR-7, SUM149PT, SUM159PT, and SUM1315 M02), xenografts were raised from 6 cell lines, while the BT549 and SUM1315M02 cell lines did not grow in mice. For the SUM159PT cell line, one of its clones (SUM159PT-luc-8A4) showed a very rapid growth rate and could give rise to a primary xenograft of the maximal size in 3.3 months, while one clone did not grow beyond the size of 4 mm^2 within the set maximum 12 months of follow-up time. The third clone of SUM159PT did not grow in the mouse mammary fat pad. Of the Hs578T cell lines, also one clone showed aggressive behavior and could give rise to the maximal tumor size in 2.5 months, whereas the other two clones could only reach 81 mm^2 in the set maximal follow-up time of 12 months. All other normal-like/claudin-low cell lines exhibited very aggressive patterns of growth and all three labeled clones of each of the cell lines gave rise to primary xenografts of the maximum size in a short time span (range 24 days to 6.5 months) (Supplementary Figure 8.2, Normal-like/claudin-low cell lines). Two clones of the WNT driven cell line DU4475[14] have also been implanted. Interestingly, these two clones showed an extremely aggressive growth behavior in the mice and the primary xenograft tumors reached the maximal allowed size in 2 weeks (Supplementary Figure 8.3). This may point to the fact that this cell line has an activated Wnt signaling pathway, which has been implicated in breast carcinogenesis[21].

Subtype specific primary xenograft tumor growth rate and tumor-related survival

The primary end-point was to determine whether xenograft growth rate and tumor-related survival in the mice was associated with the molecular subtype of the implanted cell lines. For this, we first compared the growth rates of xenografts from all cell lines across 4 molecular subtypes and then checked subtype-specific survival in the mice using Cox regression model. Our analyses revealed that the cell lines from each molecular subtype exhibited significantly different growth rates (Kruskal-Wallis Test *P* =0.01; Figure 8.2A). Importantly, cell lines of the normal-like/claudin-low molecular subtype showed a significantly higher growth rate compared with the growth rates of the cell lines of all other subtypes (Mann-Whitney Test *P* =0.001; Figure 8.2B). Additionally, the mice bearing xenografts from normal-like/claudin-low cell lines and to some extent basal-like cell lines showed poor tumor-related survival compared with mice carrying xenografts of the ERBB2-positive molecular subtype (HR: 6.89; 95% CI: 1.96-24.22; *P* =0.003 and HR: 4.30; 95% CI: 1.11-16.75; *P* =0.04, respectively) (Figure 8.2C). Surprisingly, compared with the luminal-type or basal-like cell lines the ERBB2 positive cell lines did not generate aggressive xenografts in the mice.

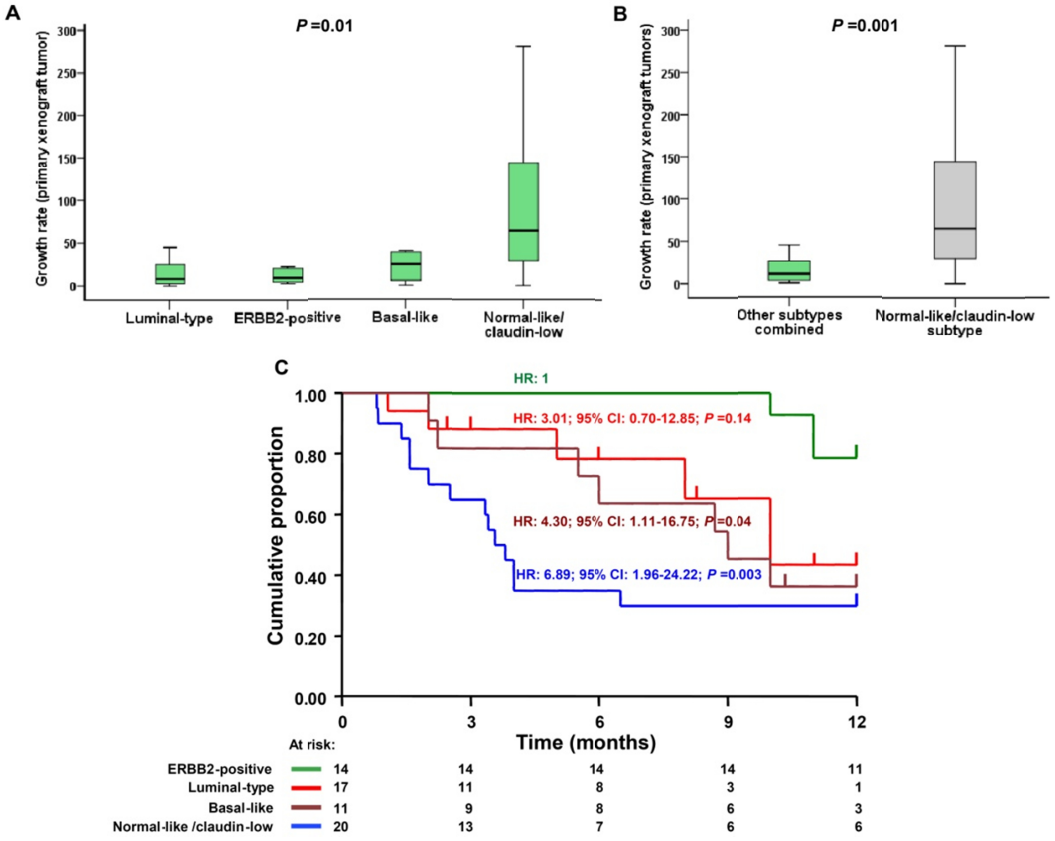


Figure 8.2 *In vivo* subtype-specific growth rate and tumor-related survival of human breast cancer cell lines. **A.** Comparison of growth rate of primary xenograft tumors from the cell lines of 4 molecular subtypes. **B.** Comparison between growth rate of primary xenograft tumors of normal-like/claudin-low cell lines and cell lines of other subtypes. Kruskal-Wallis test revealed a significant difference in the growth rate across 4 molecular subtypes while the Mann-Whitney test revealed a significant difference in the growth rates between normal-like/claudin-low cell lines and the cell lines of other three molecular subtypes combined. **C.** Xenograft tumor-related survival analysis of the cell lines across 4 molecular subtypes. Hazard ratio, 95% confidence interval, and P value for each subtype are indicated in the figure. The ERBB2-positive group was used as reference with HR set at 1. The censored animals from each subtype are also indicated in the survival curves.

Our findings that normal-like/claudin-low cell lines showed rapid growth are in line with the currently published *in vitro* growth behavior of these cell lines with doubling times for normal-like/claudin-low and basal-like cell lines being shorter compared with cell lines of the other subtypes[22]. These data suggest that the innate growth ability of these cell lines, which were propagated *in vitro* for years, did not significantly alter when implanted into the mouse mammary fat pad. Interestingly, these results also confirm clinical findings showing normal-like/claudin-low tumors to be associated with a worse prognosis. Additionally, similar to clinical specimen, these cell lines do not exhibit ER, PGR, and ERBB2 expression (triple-negative)[23] and a rather high expression of mesenchymal markers such as $CD44^{high}/CD24^{low}$ (known breast cancer stem cell markers)[10] (Hollestelle, submitted paper) has been reported. In contrast to the clinical observations[24,25], none of the ERBB2-positive cell lines exhibited aggressive growth and poor survival in mice (Figure 8.2B and C). This may be, on the one hand, due to the small number of ERBB2-positive cell lines that successfully were implanted in the mice. On the other hand, ERBB2-positive cell lines showed a slower growth rate in *in vitro* culture conditions compared with basal-like and normal-like/claudin-low cell lines[14].

Subtype specific metastatic potential and distant organ specificity

Besides variable growth patterns associated with the cell lines, the propensity of these cell lines to colonies common distant sites in mice were determined, as well as whether this behavior was molecular subtype related. For this, we only focused on those mice in which primary xenograft tumors were able to reach the maximal allowed size ($225mm^2$) because smaller sized xenograft tumors did not metastasize to the distant sites. As expected, the most common sites of relapse were the adjacent lymph nodes (73%). Second in line to lymph nodes were lungs (43%). Liver and uterus/ovaries (27% each) were the third most frequently colonized distant organs by the cell lines (Figure 8.3 and Supplementary Table 8.1).

Relapses to spleen, bone, and brain were also observed though less frequently compared with those to lymph nodes and lungs. Additionally, occasional relapses to some uncommon sites with regard to breast cancer, such as to the kidney and pancreas were also observed in the mice (Supplementary Table 8.1). It should however be noted that the majority of these micrometastases were only detectable by the bioluminescence and could not be confirmed histologically except for the metastases to lymph nodes and lungs (Figure 8.4).

It has been previously observed that the intrinsic molecular subtypes show preferential sites of distant relapses[6,7]; and also intravenously injected breast cancer cells show organ-specific relapse[26-28]. Unfortunately, in mice, these subtype-specific preferences in the cell lines of 4 molecular subtypes could not be reproduced. One explanation could be that our panel of cell lines is small and therefore lacks sufficient power to observe subtype-specific sites of relapse in the mice. Another possible reason could be that the xenografts raised in the mammary fat pads are more prone to shed off cells directly into the body lumen, which subsequently can metastasize non-preferentially to various body organs.

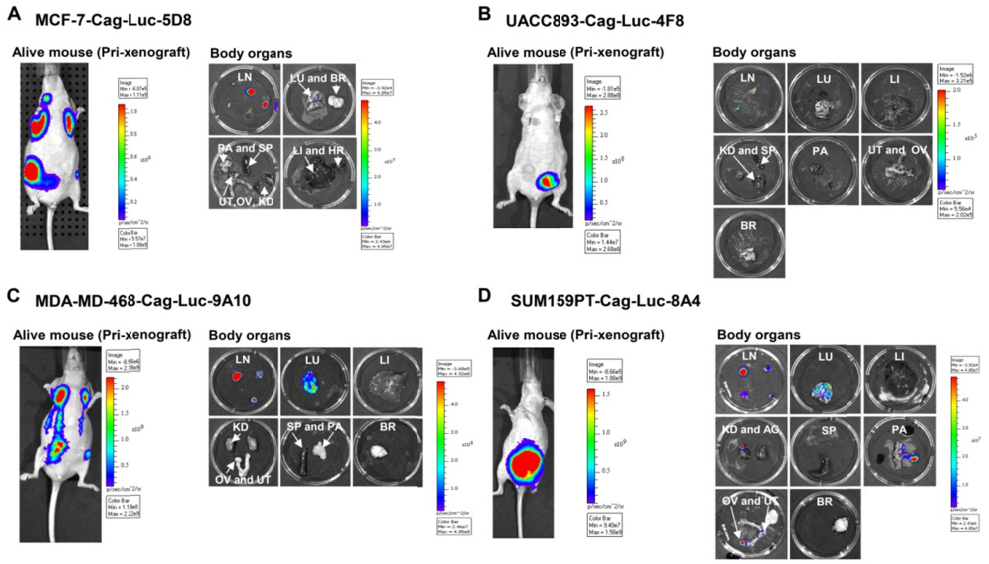


Figure 8.3 Bioluminescence imaging of the intact animals as well as organs showing micrometastases of representative cell line xenografts from four molecular subtypes. A. Primary xenograft tumor and metastatic lesions on distant sites from a MCF-7-Cag-Luc-5D8 luciferase-labeled clone, a representative picture of the luminal-subtype of cell lines, B. Primary xenograft tumor and metastatic lesions on distant sites from a UACC893-Cag-Luc-4F8 luciferase-labeled cell line, a representative picture of the ERBB2-positive subtype of cell lines, C. Primary xenograft tumor and metastatic lesions on distant sites from a MDA-MB-468-Cag-Luc-9A10 luciferase-labeled clone, a representative picture of the basal-like subtype of cell lines, and D. Primary xenograft tumor and metastatic lesions on distant sites from a SUM159PT-Cag-Luc-8A4 luciferase-labeled clone, a representative picture of the normal-like/claudin-low subtype of cell lines. LN, lymph-nodes; LU, lungs; LI, liver; PA, pancreas; KI, kidneys; AG, adrenal glands; SP, spleen; OV, ovaries; UT, uterus; BR, brain.

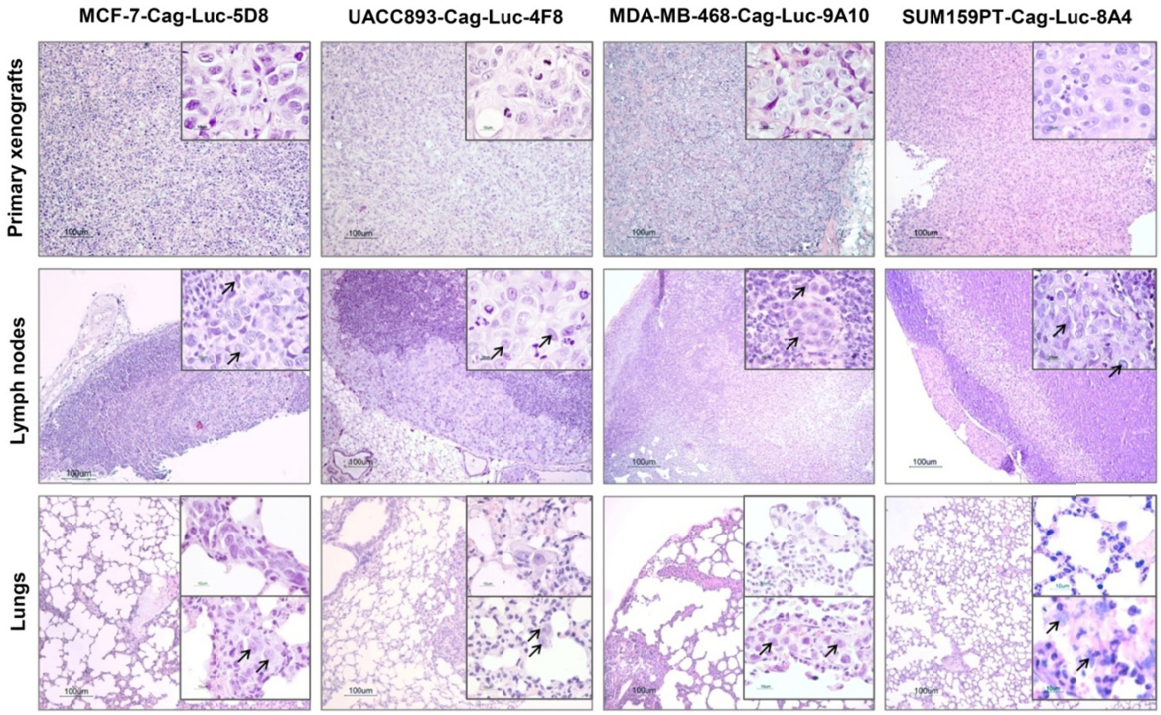


Figure 8.4 A representative figure showing hematoxylin and eosin staining of the primary xenografts and metastases to lymph nodes and lungs. Staining of four representative luciferase-labeled clones from four molecular subtypes is shown. Arrows in the inserts indicate metastatic tumor cells in the tissue at 10 time higher magnification.

Moreover, frequent relapses to the lungs and rare relapses to the bone in mice suggest that cells disengaging from orthotopically raised tumors may take a long time to establish metastases to the organs located at a far distance from the primary site. To follow metastases to these far located organs in mice would therefore be challenging due to age limitations of the mice and fast appearance of the recurrent tumor. Intra-tail and intra-cordial injection of these cell lines may therefore shed more light on the organ specific metastatic potential of these cell lines.

Breast cancer cell line xenografts as preclinical models

Next, we investigated whether this panel of BC cell lines can robustly be used as a xenograft preclinical mouse model. For this, the robustness of the growth rate of three representative luciferase labeled cell lines (MDA-MB-468-Cag-Luc-9A10, MDA-MB-231-Cag-Luc-5D6, and SUM159PT-Cag-Luc-8A4) was assessed (Supplementary Figure 8.4). Slices of 2 mm^3 from primary xenograft tumors were serially implanted into three additional recipient mice. As expected, the take rate of the daughter xenograft tumors was 100%. Importantly, the growth rate of three serially implanted xenograft tumors from each cell line was uniform as concluded from the small error bars and rapid compared with their corresponding parental xenograft tumors generated directly from the cultured cells (Figure 8.5). These data show that the BC cell line xenograft model could potentially be an attractive preclinical model for drug screening. Importantly, these xenografts grew in a practically feasible time frame and were accompanied with mouse stroma, which has been suggested to have a critical effect on the sensitively of the tested drugs[29]. Moreover, as suggested previously[15], the group of cell lines with the same molecular subtype can be used as a system rather than individually for better and more robust screening of the drugs.

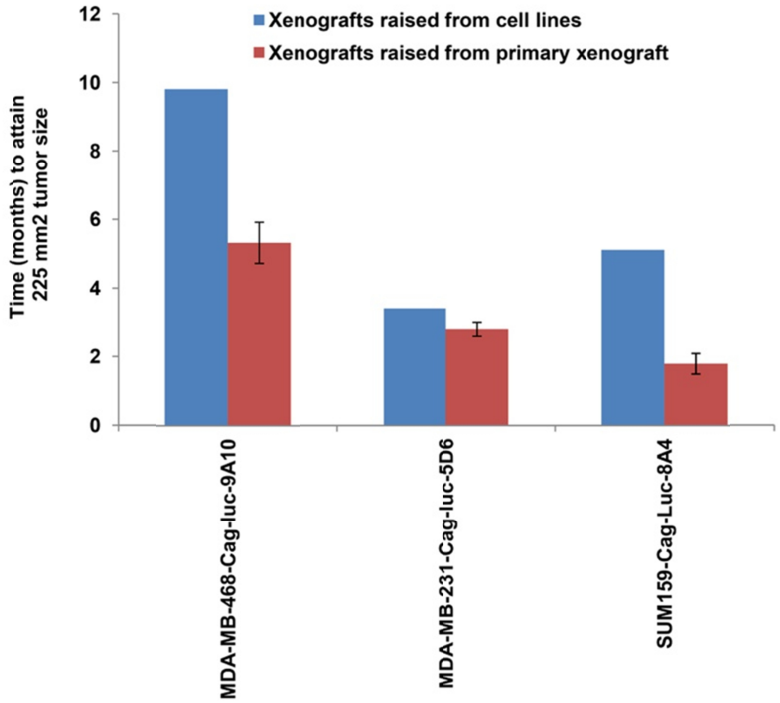


Figure 8.5 Comparison between growth rates of the xenograft tumors generated from cell lines and from serially implanted 2 mm^3 tumor cubes of the primary xenografts from the same cell lines.

CONCLUSIONS

In this study we show that human breast cancer cell lines, which have been propagated on plastic for years, have retained their innate abilities to exploit the (orthotopic) microenvironment of the implantation site and can give rise to xenografts when implanted into the mammary fat pad of the mice. The most common sites colonized by all types of cell lines in mice are lymph nodes, lungs, liver and ovaries/uterus, whereas some other sites of relapse such as kidneys, bone, brain and pancreas were also rarely observed. Importantly, xenografts from breast cancer cell lines of the normal-like/claudin-low, and to a lesser extent of the basal-like subtype, showed a more aggressive growth behavior and worse tumor-related survival of the mice compared with xenografts of cell lines from the other molecular subtypes. Finally, the serially generated xenografts in the recipient mice grew reproducibly and robustly and can serve as diverse and practically feasible preclinical models to screen drugs *in vivo* in the presence of mouse stroma.

ACKNOWLEDGEMENTS

This study was supported financially by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO). We appreciated the help provided by the colleagues of the animal facility. We acknowledge Prof. B. van der Hoest from Genetics Department of Erasmus University Medical Center for his generous donation of HEK293T cells and plasmid with luciferase construct. We are also thankful to Roel Janssen from the Department of Genetics, Erasmus University Medical Center for his help in generating lentivirus particles and cell line transduction.

CONFLICT OF INTEREST STATEMENT

All authors declared that they have no competing interest.

SUPPLEMENTARY DATA

Supplementary Table 8.1 and Supplementary Figures 8.1, 8.2, 8.3 and 8.4 are available on:
https://www.dropbox.com/sh/4fwrk0lit91dsp5/WP1sZYZAQw?m

REFERENCES

1. Weigelt, B. & Reis-Filho, J.S. Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat Rev Clin Oncol 6, 718-30 (2009).
2. Weigelt, B. et al. Refinement of breast cancer classification by molecular characterization of histological special types. J Pathol 216, 141-50 (2008).
3. Perou, C.M. et al. Molecular portraits of human breast tumours. Nature 406, 747-52 (2000).

4. Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100, 8418-23 (2003).
5. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98, 10869-74 (2001).
6. Smid, M. et al. Subtypes of breast cancer show preferential site of relapse. Cancer Res 68, 3108-14 (2008).
7. Kennecke, H. et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol 28, 3271-7 (2010).
8. Smid, M. et al. Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. Breast Cancer Res Treat 128, 23-30 (2011).
9. Farmer, P. et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 24, 4660-71 (2005).
10. Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 12, R68 (2010).
11. Lehmann, B.D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121, 2750-67 (2011).
12. Garnett, M.J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483, 570-5 (2012).
13. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603-7 (2012).
14. Hollestelle, A. et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res Treat 121, 53-64 (2010).
15. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515-27 (2006).
16. Dull, T. et al. A third-generation lentivirus vector with a conditional packaging system. J Virol 72, 8463-71 (1998).
17. Yee, J.K. et al. A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. Proc Natl Acad Sci U S A 91, 9564-8 (1994).
18. Boussif, O. et al. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A 92, 7297-301 (1995).
19. Alexopoulou, A.N., Couchman, J.R. & Whiteford, J.R. The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. BMC Cell Biol 9, 2 (2008).
20. Nguyen, A.T., Dow, A.C., Kupiec-Weglinski, J., Busuttil, R.W. & Lipshutz, G.S. Evaluation of gene promoters for liver expression by hydrodynamic gene transfer. J Surg Res 148, 60-6 (2008).
21. Schlange, T., Matsuda, Y., Lienhard, S., Huber, A. & Hynes, N.E. Autocrine WNT signaling contributes to breast cancer cell proliferation via the canonical WNT pathway and EGFR transactivation. Breast Cancer Res 9, R63 (2007).

22. Heiser, L.M. et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A 109, 2724-9 (2012).
23. Hennessy, B.T. et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res 69, 4116-24 (2009).
24. Buzdar, A.U. et al. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in human epidermal growth factor receptor 2-positive operable breast cancer. J Clin Oncol 23, 3676-85 (2005).
25. Gonzalez-Angulo, A.M. et al. High risk of recurrence for patients with breast cancer who have human epidermal growth factor receptor 2-positive, node-negative tumors 1 cm or smaller. J Clin Oncol 27, 5700-6 (2009).
26. Bos, P.D. et al. Genes that mediate breast cancer metastasis to the brain. Nature 459, 1005-9 (2009).
27. Minn, A.J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518-24 (2005).
28. Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 3, 537-49 (2003).
29. Straussman, R. et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. Nature 487, 500-4 (2012).

Chapter 9

General discussion

GENERAL DISCUSSION

Breast cancer is the most common cancer among women world-wide. Like many other diseases, breast cancer is also a genetic disease that emerges due to accumulation of mutations in biologically important genes. The mutated genes are either inherited from parents (hereditary breast cancer) or the mutations are acquired de novo (sporadic breast cancer) during the course of life. Hereditary breast cancer mainly includes familial breast cancer in which gene mutations cluster into a family and therefore affected individuals in the family are at higher risk of developing breast cancer compared to the general population. The genetic basis of inherited breast cancer has been investigated for the past three decades and has been the subject of several recent discoveries. Three reasonably well-defined groups of breast cancer susceptibility genes have become apparent: rare high-risk and moderate-risk genes and common low-risk loci. The rare high-risk genes were discovered through genetic linkage mapping and positional cloning[1-8] and the mutations in these genes confer a risk of approximately 30-60% of developing breast cancer by the age of 60, compared to 3% in the general population[9]. The moderate-risk genes were discovered through a direct interrogation of the candidate genes in genetically enriched cases and controls[10-14]. Mutations in these genes are relatively rare and they confer a risk of about 20-40% by the age of 60[15]. The low-risk breast cancer alleles have been discovered through genome-wide association studies (GWAS) studying hundreds of thousands of single nucleotide polymorphisms (SNPs) in large populations of breast cancer cases and controls[16,17]. The low-risk alleles are very common (population frequency 5-50%) and they confer a modest risk of 10-20% by the age of 60. Altogether, these three groups of breast cancer susceptibility genes/loci can currently only explain about 35% of familial breast cancer cases, whereas 65% of the familial breast cancer risk still remains to be elucidated.

Based on the extent to which breast cancer risk has been estimated in twin and sibling studies, it is expected that a considerable part of the familial breast cancer risk is genetic, which indicates that a large number of risk loci and/or SNPs are still to be discovered. In line with this, Park and colleagues estimated that if SNPs explain all the unexplained risk approximately 60-70 SNPs with effects similar in size to those already known are likely to be around[18]. Next to the discovery of novel risk factors, the complete clinical significance and the biological function of the newly discovered low-risk breast cancer susceptibility loci is still to be understood. Furthermore, a wide variety of well-characterized *in vivo* models to study genetic and biological factors are lacking, which hampers the understanding of their role in disease progression.

In the first experimental chapter of the thesis, using candidate gene approach we tried to discover a new gene or variant in *CHEK2* mutation carrier breast cancer families. This putative gene or variant would in our view not only explain a part of yet unexplained familial breast cancer risk but also act as a partner of the *CHEK2* gene in a polygenic setting. Furthermore, in the two follow-up chapters we investigated in cell lines and in primary breast tumors the biological mechanisms and additional clinical significance associated with the first series of common low-risk breast cancer loci previously identified via GWAS. In the **second part** of the thesis, we investigated breast cancer

subtype-specific biological entities such as genes and micro-RNAs (miRNAs) using either primary breast tumors or *in vitro* propagated breast cancer cell lines. **Lastly**, we developed breast cancer cell line xenograft mice models to determine if subtype-related clinical behavior of breast cancer is recapitulated in these models and also to functionally explore the biology of the clinically relevant factors as well as to investigate interference of currently existing as well as novel targeted drugs.

Discovering the polygenic partner of CHEK2 for breast cancer susceptibility

Our group has previously shown that the *CHEK2* 1100delC variant confers a moderate risk of approximately 2-fold in breast cancer families with no BRCA1 and BRCA2 mutations. The inheritance pattern of breast cancer in *CHEK2* 1100delC carrier families suggested that these families are of as much high-risk families as BRCA1 and BRCA2 families[10,19] E.g. a meta-analysis of unselected 26,000 *CHEK2* 1100delC mutation carrier cases and 27,000 controls have estimated an odd ratio (OR) of 4.8[20]. Furthermore, it was shown that the *CHEK2* 1100delC variant alone conferred only an about 2-fold increased risk of breast cancer and the variant did not completely segregate with breast cancer in these families[10,19]. These observations have led to the hypothesis that *CHEK2* 1100delC co-segregates with an as-yet unknown gene, which may function with the *CHEK2* 1100delC variant in a polygenic setting in *CHEK2* 1100delC carrier breast cancer families[21,22]. Looking at the OR of 4.8 determined in the meta-analyses and considering that *CHEK2* 1100delC may operate in a multiplicative polygenic model, either one putative gene 'X' with a moderate-risk susceptibility comparable to the *CHEK2* 1100delC variant or two putative alleles 'X and Y' both with a low-risk susceptibility comparable to common variants are expected in *CHEK2* 1100delC mutation carrier families (Figure 9.1). To reveal this *CHEK2* 1100delC polygenic partner, we previously evaluated mutations in a couple of candidate genes for their possible involvement in the polygenic CHEK2 cancer model. The evaluation of the three most common Dutch *MUTYH* mutations, MUTYH Y179C, G396D, and P405L, in 70 Hereditary Breast Colorectal Cancer (HBCC) families, which were also evaluated for the *CHEK2* 1100delC variant, revealed only one of the 70 HBCC families carriers having mutations in both *CHEK2* and *MUTYH* genes. This finding suggested that the *MUTYH* mutations do not play a major role in the polygenic *CHEK2* cancer model[23]. Secondly, screening of a SNP in the promoter of the *MDM2* gene (*MDM2* SNP209) in 20 CHEK2 1100delC familial breast cancer cases and 126 geographically-matched controls also showed no evidence of *MDM2* SNP209 to be of major importance in the polygenic *CHEK2* cancer model[23] (Figure 9.1). In chapter 3 of this thesis, we evaluated the *MSH6* gene as a possible polygenic partner of *CHEK2* in breast cancer by performing mutation analysis of the *MSH6* gene in breast cancer families. Though, while we did not find any mutations in the coding part of the *MSH6* gene in breast cancer families, we did observe some rare variants having a population frequency of less than 1% to be associated with increased breast cancer risk.

Our choice of considering *MSH6* as a likely candidate of being the polygenic partner *CHEK2* was based on previous observations that germline mutations in all mismatch repair genes including *MSH6* are associated with colorectal cancer[24-26]. Also, it was shown previously by our group that *CHEK2* 1100delC also confers colorectal cancer risk in breast cancer families with the HBCC

phenotype. Importantly, both *CHEK2* and *MSH6* have been implicated in *lynch syndrome*[27,28]. We therefore hypothesized that the mutations in the *MSH6* gene might be involved in polygenic *CHEK2* 1100delC breast cancer susceptibility. Sequencing analysis of the *MSH6* in 68 HBCC breast cancer families revealed no obvious mutations, indicating that the *MSH6* gene is not a major player in the polygenic CHEK2 cancer model. Interestingly however, some rare variants spanning the entire *MSH6* coding sequence and surrounding exon-intron boundaries were identified. The combined frequencies of these variants were significantly increased in HBCC families compared to population matched-controls [6/136 (11.8%) vs. 5/306 (~1.5%), *P* <0.00001]. In addition, these variants were also found more often in a cohort of 68 matched non-HBCC families, suggesting a general association of these variants with breast cancer susceptibility. None of these rare variants are clear pathogenic and the underlying disease mechanism of the oncogenic effect is therefore unknown. However, as suggested by a recent study from Encylopedia of DNA Elements (ENCODE) Project consortium[29] one may speculate that these rare variants may change binding affinity of the transcription factors or other chromatin remodelling factors. Furthermore, they may alter splicing machinery efficiency or specificity or may change miRNA target sequences. These effects subsequently would lead to an altered expression of either the *MSH6* gene itself or another protein-coding gene located on the same linkage disequilibrium (LD) block. Expression analysis of the *MSH6* gene in relation to these rare variants, as we have done for low-risk SNPs in chapter 4 and 5, may provide clues to whether a putative disease mechanism involves expression modulation of the *MSH6* gene.

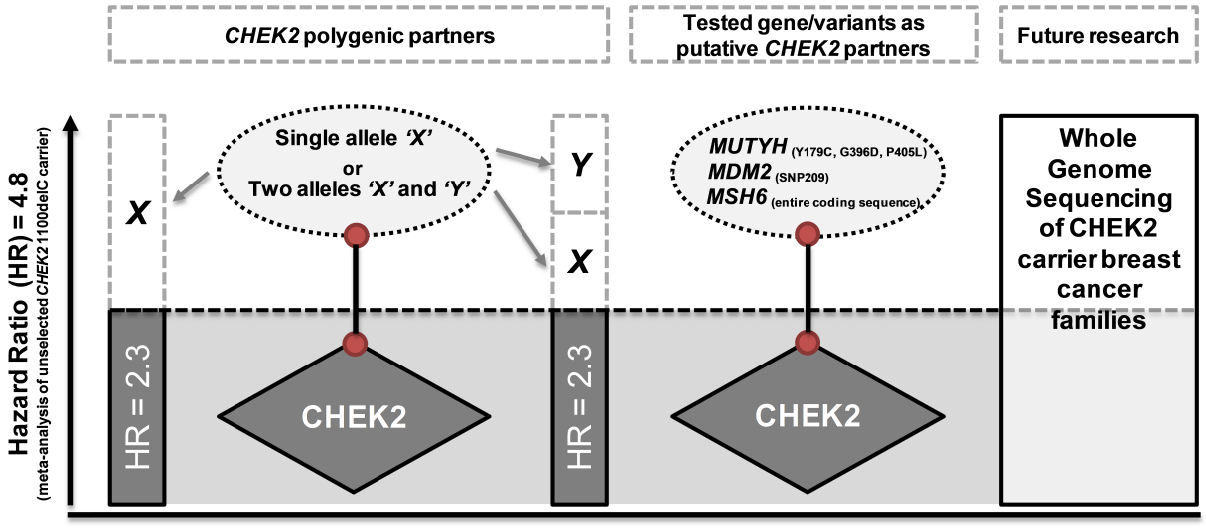


Figure 9.1 Schematic representation of the CHEK2 polygenic model for breast cancer susceptibility, currently tested candidate genes and future recommendation.

As target gene sequencing efforts by a candidate gene approach have not yielded any gene or alleles to be involved in a CHEK2 polygenic model for breast cancer and given that whole genome sequencing is becoming cost-effective over time, in future studies would therefore be promising to consider whole exome or whole genome sequencing to explore putative gene or variants that possibly interact with the *CHEK2* 1100delC variant. Conceivably, individuals from *CHEK2* 1100delC

carrier breast cancer families are expected to be genetically enriched for this putative gene/allele. These families would therefore serve as the most promising candidates to discover a putative gene/allele by whole exome or whole genome sequencing and to provide evidence that this putative gene/allele indeed acts in concert with the *CHEK2* 1100delC variant. The discovery of such gene would likely explain the remaining magnitude of unexplained risk of developing breast cancer in high-risk *CHEK2* 1100delC families and thus potentially be paving the way for genetic counseling of high risk cases in these families. In this regard our lab has anticipated the potential of whole exome genome sequencing and is already on the way towards applying this approach to search for the CHEK2 polygenic partner for breast cancer susceptibility.

Common low-risk breast cancer loci and their biological and clinical significance

In the past two decades, the high- and moderate-risk breast cancer susceptibility genes have been extensively studied. The underlying biology and disease mechanism associated with these genes have been comprehensively explored. Accordingly, different strategies were developed to design preventative therapeutics to fight breast cancer associated with these genes. A prime example is the clinical development of PARP inhibitors in oncology, which became only possible due to a thorough fundamental biological knowledge in BRCA1/BRCA2-deficient tumors[30]. Likewise, it is of great importance to discover disease mechanisms and related biological pathways that underlie the emerging common low-risk breast cancer loci that have been discovered through GWAS using SNPs in breast cancer cases and controls. The thorough mechanistic understanding of these loci would allow the design of effective drugs to target breast cancer associated with these loci in the clinic. The fact that the majority of these low-risk SNPs are mapped to non-coding regions of the genome has made it less obvious how such a locus would functionally contribute to breast carcinogenesis. In 2007 the Breast Cancer Association Consortium (BCAC) postulated two hypotheses[6] about disease mechanisms through which low-risk loci may confer their increased breast cancer risk:

1. The disease-associated SNPs tagging low-risk breast cancer susceptibility loci are linked with yet unknown causal variants in protein coding genes located in the same LD block (Figure 9.2A).

2. The low-risk breast cancer susceptibility loci may function through expression modulation of nearby located genes and such expression changes are causally related to breast carcinogenesis (Figure 9.2B).

To explore the first hypothesis that risk-associated SNPs are linked with causal variants in nearby genes, BCAC performed deep sequencing of the genes containing these loci or all the genes located in the same LD block. Such sequencing efforts, however, did not reveal any pathogenic variants correlated with the minor allele of the identified loci[16]. So the first hypothesis was therefore considered unlikely to be of great significance.

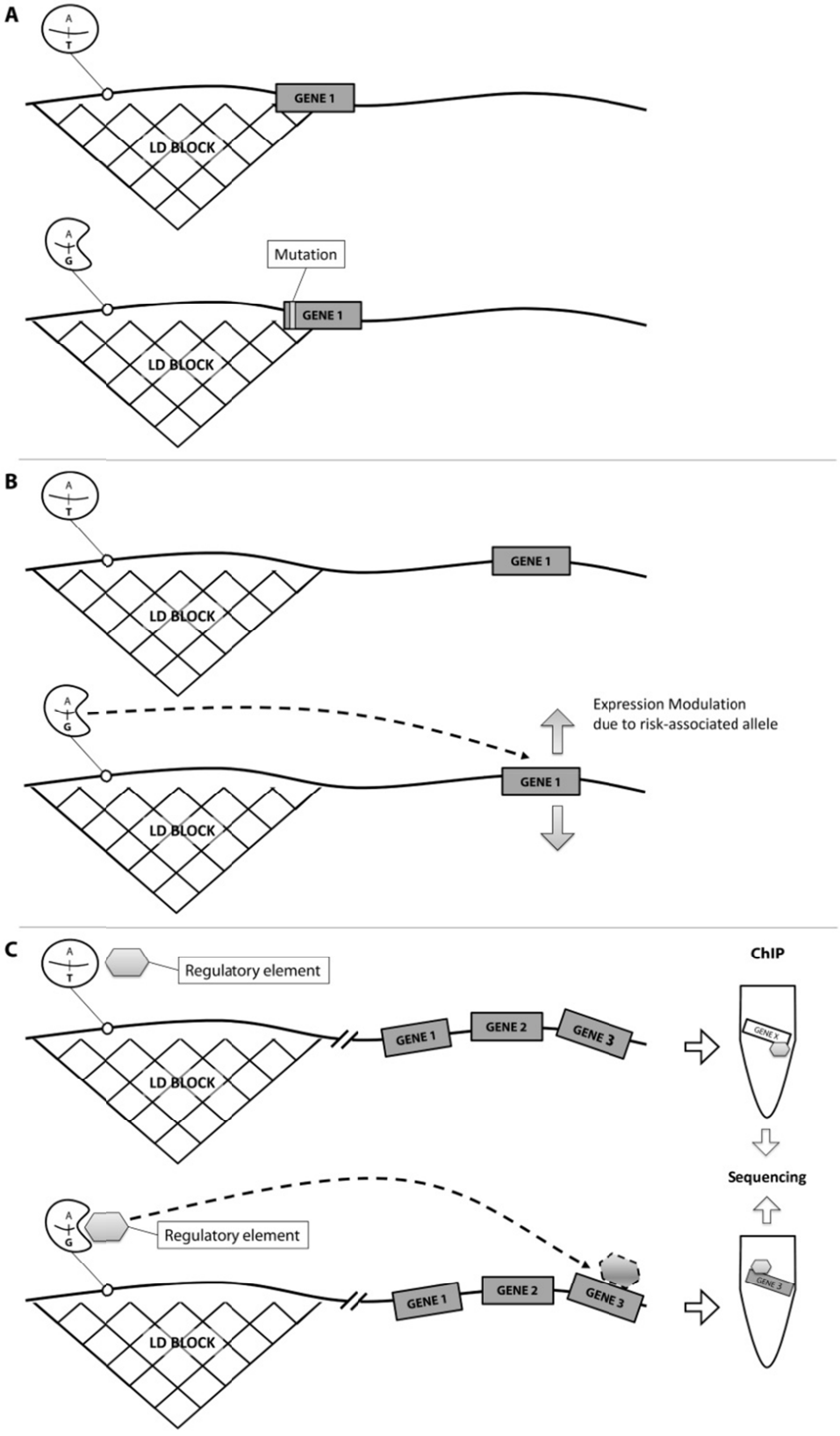


Figure 9.2 Schematic diagram showing possible disease mechanisms of common low-risk breast cancer SNPs. **A.** Correlation of risk-associated SNP with a causal mutation in the gene located on the same linkage disequilibrium (LD) block, **B.** Association of risk-associated SNP with expression modulation of the nearest gene (two BCAC hypotheses) and **C.** Risk-associated SNP interacts with a causal gene via long range chromatic loop by altering binding affinity of the regulatory elements.

In **chapter 4 and 5** of this thesis, we evaluated the second hypothesis, i.e., the low-risk breast cancer loci are associated with expression modulation of the nearest genes. For this, we correlated in 1,401 primary breast cancer patients mRNA expression levels of the genes nearest to the genotyped risk-conferring SNPs. Among the seven loci explored (FGFR2-rs2981282, LSP1-rs3817198, MAP3K1-rs889312, TOX3-rs3803662, H19-rs2107425, 2q35-rs13387042, and 8q24-rs13281615), we found that only the risk-associated allele of rs3803662 SNP around the *TOX3* gene was statistically significantly associated with lower mRNA expression of the *TOX3* gene. This significant association of the rs3803662 SNP with lower expression of *TOX3* suggests a tumor suppressor role of this gene in breast carcinogenesis. Of note, expression of *TOX3*, a high mobility group box nuclear protein involved in calcium dependent transcription and structural modification of chromatin, has previously been implicated by us in bone metastasis in the breast cancer patients[31]. Our findings that the risk-associated locus (rs3803660) nearby *TOX3* at 16q12.1 was indeed associated with lower expression of *TOX3* in breast cancer cells and that the *TOX3* gene is a new tumor suppressor gene in breast cancer have been recently reproduced by Cowper-Sal Lari and colleagues[32]. They annotated breast cancer-risk associated SNPs in the non-coding regions of the genome in breast cancer cells and provided functional evidence that these SNPs are enriched in binding sites for FOXA1 and ESR1 transcription factors and carry more frequently H3K4me1 chromatin marks in cancer in a cell-type specific manner. They further showed that the most significantly associated SNP rs4784227 with breast cancer loci at 16q12.1 regulates the *TOX3* expression in an allele-specific manner. This SNP rs4784227 is in significant LD with rs3803660 SNP ($r = 0.864$, $D' = 1$), which was originally shown by BCAC to be associated with increased breast cancer risk to the same locus and we showed it to be associated with lower expression of the *TOX3* gene. The risk-associated allele of rs4784227 [T] allows an increased binding affinity for *FOXA1* transcription factor on the enhancer, which tempers *TOX3* transcription via a chromatin loop interacting with *TOX3* promoter. Additionally, Cowper-Sal Lari and colleagues silenced *TOX3* in ZR-75-1 breast cancer cell lines to determine whether *TOX3* acts as a tumor suppressor gene. Indeed, knockdown of the *TOX3* gene resulted in enhanced proliferation of ZR-75-1 cells providing functional evidence that *TOX3* indeed has a tumor suppressor role in breast carcinogenesis.

A similar mechanism has been reported for rs2981582 and rs13281615 breast cancer risk-associated SNPs, which are located in intron 2 of *FGFR2* and in the 8q24 locus, respectively[33,34]. The rs2981582 SNP was shown to be a causal variant by itself, as it regulates increased *FGFR2* expression by altering the binding affinity of transcription factors Oct1/Runx2 and C/EBPβ on the genomic region containing this SNP[33]. Likewise, risk variant of rs13281615 on 8q24 was shown to differentially regulate the expression level of the *MYC* gene through physical interaction with tissue-specific enhancers from the 8q24 region containing the SNP[34]. This biological mechanism for a breast cancer risk-associated SNPs, which involves the disruption of transription factor binding at a distal enhancer and the subsequent modulation of the gene expression as reported for *TOX3*, *FGFR2*, and *MYC* genes, appears to be a common mechanism of action. In general, this mechanism involves immunoprecipitation of the regulatory elements suspected to bind to genomic region containing risk-associated variant and also interact with the target gene. The captured DNA in the

precipitates is subsequently sequenced to determine the exact genomic parts invoveld in the interaction (Figure 9.2C). Such mechanism has recently gained support on a global scale as the findings published by the ENCODE consortium also provided support for enrichment of regulatory elements at SNPs linked by the GWAS to various diseases. It is important to mention however that we in our cohort of 1,401 primary breast cancers do not find support for *FGFR2* and *MYC* expression changes in relation to the risk-associated SNPs, nor for the other 4 loci investigated, indicating a need for replication of our studies in even larger cohorts of tumors. Alternatively, these SNPs are not linked to expression changes of the indicated genes, or the expression changes are not of relevance in the tumor tissue but in the normal tissue from which the tumor is derived.

The number of GWAS-identified breast cancer risk-associated SNPs is large and this number is still growing. Now it is the time for the scientific community to be more focused on the post-GWAS phase of human genetics. A comprehensive set of priniciples has already been proposed for the post-GWAS functional characterizations of low-risk cancer loci[35]. However, the major complication associated with the functional characterization of these SNPs is that they themselves are not likely causal but are associated with hundreds or even a thousand of linked SNPs in the genome of which one is likely to be causal. Each of these linked SNPs may have multiple potential target genes through a number of potential interactions mediated by chromatin loops. This makes it difficult to discriminate causal SNPs from a SNP that is only associated with the increased cancer risk. Our current knowlegde calls for a systematic approach that can tell them apart effectively. In this regard, the methods recently developed and applied by Cowper-Sal Lari and colleagues[32] seem promising as they first reveal the most promising transcription factors, histone modifications or other annotations that are specific to the genomic regions containing the risk-associated SNP. Secondly, they identify for experimental follow-up studies from thousands of associated SNPs a candidate SNP that affects the interaction of annotated factors with the chromatin. This approach would eventually pave the way leading to an integrated post-GWAS characterization of low-risk cancer susceptibilty loci.

The second question that we addressed in **Chapter 5** is whether these low-risk loci besides being associated with breast cancer onset also predict progression of the disease in breast cancer patients. Our study identified only one SNP that is rs2107425, near *H19* gene, which was associated with shorter metastasis free-survival in patients with primary breast cancer. A recent BCAC study also investigated 11 SNPs that were associated with breast cancer risk and 62 other SNPs, which were suggested as breast cancer risk SNPs by a GWAS or identified as candidate SNPs from individual studies, for their prognostic relevance[36]. The 11 SNPs investigated in this study also include the 7 SNPs that we tested in our study. One of the 11 SNPs, rs3803662 (*TOX3*) and none of the 62 candidate/GWAS SNPs were associated with overall survival (OS) and/or breast cancer-related survival (BCS)[36]. The genotypic-specific survival for rs3803662 suggested a recessive mode of action [hazard ratio (HR) of rare homozygous carriers = 1.21; 95% CI: 1.09–1.35, P =0.0002 and HR = 1.29; 95% CI: 1.12–1.47, P =0.0003 for OS and BCS, respectively][36]. However, we were not able to detect a prognostic association of rs3803662 (*TOX3*) SNP in our relatively small cohort of 1,290 lymph node-negative patients. The major difference between our cohort and the BCAC cohort is

that our SNP genotyping data have been derived from tumor DNA, which mainly comprised of sporadic tumors, whereas BCAC SNP genotypes were derived from blood-based and thus germline DNA. Also, a significant percentage of patients in BCAC study were treated with adjuvant systemic therapy, which may have confounding effect on the progression of the disease. The data from both studies however suggest that low-risk breast cancer loci are mostly, if not solely, related to the onset of the disease and it is less likely that they affect progression of the disease in patients.

TWIST1 expression and its clinical and biological significance

In **chapter 6** of this thesis, we determined mRNA expression levels of the *TWIST1* gene as a marker of progression of primary breast cancer particularly in untreated lymph node-negative patients. Importantly, we found that high *TWIST1* mRNA expression was significantly associated with poor prognosis only in the estrogen receptor(ER)- positive lymph node-negative subgroup of patients. These findings are in line with two previous studies showing that TWIST1 protein expression determined by IHC predicts poor prognosis in ER-positive breast cancer[37,38]. In addition to that, we observed particularly in ER-positive tumors that *TWIST1* mRNA expression levels are significantly higher in stromal-rich compared with stromal-poor tumors, implying that the stroma of the ER-positive tumor tissue could be an important factor involved in *TWIST1*-related breast cancer progression. Interestingly, a connection between *TWIST1* expression and tumor stroma in the breast tissue has also recently been demonstrated by Roman-Perez and colleagues[39], who measured global gene expression patterns of 72 breast cancer tissues and identified two distinct subtypes, active and inactive, in the cancer-adjacent extra-tumoral microenvironment of breast tissues. The data showed that the *TWIST1* was highly expressed along with other stroma-associated genes such as *VIM, ADAMTS2, COL4A2, COL4A1, ITGA7*, and *ITGB1* in the active subtype. Fitting with this, we showed in microdissected tumor tissues that *TWIST1* gene at the mRNA level is almost exclusively expressed in the stromal compartment of the tissue. Also, our *in silico* pathway analysis using *TWIST1* co-expressed genes showed an enrichment of extracellular matrix (ECM)-related gene ontology (GO) terms and biological pathways. Importantly, similar to mRNA expression, TWIST1 protein expression was higher, but not exclusively, expressed in the stroma of the ER-positive tumor tissues compared with that of ER-negative tissues. Our results suggest a biological relationship among stromal components, *TWIST1* and prognosis specifically in ER-positive breast tumor tissue. Historically, TWIST1 has repeatedly been linked to tumor invasion and metastasis[40]. A significant body of evidence demonstrated that TWIST1 induces metastasis by promoting an epithelial-to-mesenchymal transition (EMT) in various cancer types including breast cancer[41,42]. The TWIST1-mediated EMT process in epithelial cells generally involves transcriptional repression of the *E-cadherin* gene by TWIST1 expression, which leads to loosening of cell-cell adhesion and activation of mesenchymal markers. Consequently, malignant epithelial cells become more motile and metastatic. In breast cancer, it has been observed that TWIST1 induces EMT preferentially in basal-like breast tumors[43], which are typically ER-negative cancers. Consistent with this, the majority of ER-negative cell lines display EMT features[44]. In our study, we observed that *TWIST1* mRNA expression levels in ER-positive breast cancers was positively correlated with the known EMT

markers such as *ZEB1*, *ZEB2, Snail2*, and alpha-smooth muscle actin (*α-SMA*) suggesting that the EMT might occur in ER-positive cancer as well. However, we did not observe a clear negative correlation between *TWIST1* and *E-cadherin* expression, another well-known hallmark of EMT, to support that *TWIST1* may functions via EMT in ER-positive tumors.

In view of previous literature that TWIST1 induces metastasis in various cancers and in the light of our current findings that predominant stromal *TWIST1* expression is associated with poor prognosis in ER-positive breast cancer, we envision two scenarios of TWIST1-mediated metastasis in breast cancers. The first scenario is the well-characterized *'TWIST1-induced EMT scenario'* in which TWIST1 induces metastasis via EMT involving a clear expression modulation of the EMT markers such as downregulation of E-cadherin and catenins expression and upregulation of vimentin, fibronectin and *α-SMA* expression (Figure 9.3A). During the EMT, epithelial cells undergo morphological changes by losing their epithelial morphology and acquiring mesenchymal characteristics and thereby becoming more motile, allowing tumor cells to escape through the basement membrane into the blood circulation eventually giving rise to metastases in the distant organs. The second scenario is *'TWIST1-induced stroma activation scenario'* in which TWIST1 causes structural remodeling of stromal components such as vimentin, integrins, collagens and the ADAM family of matrix metalloproteinase (MMP) enzymes (*ADAMTS)* in tumor tissue (Figure 9.3B).

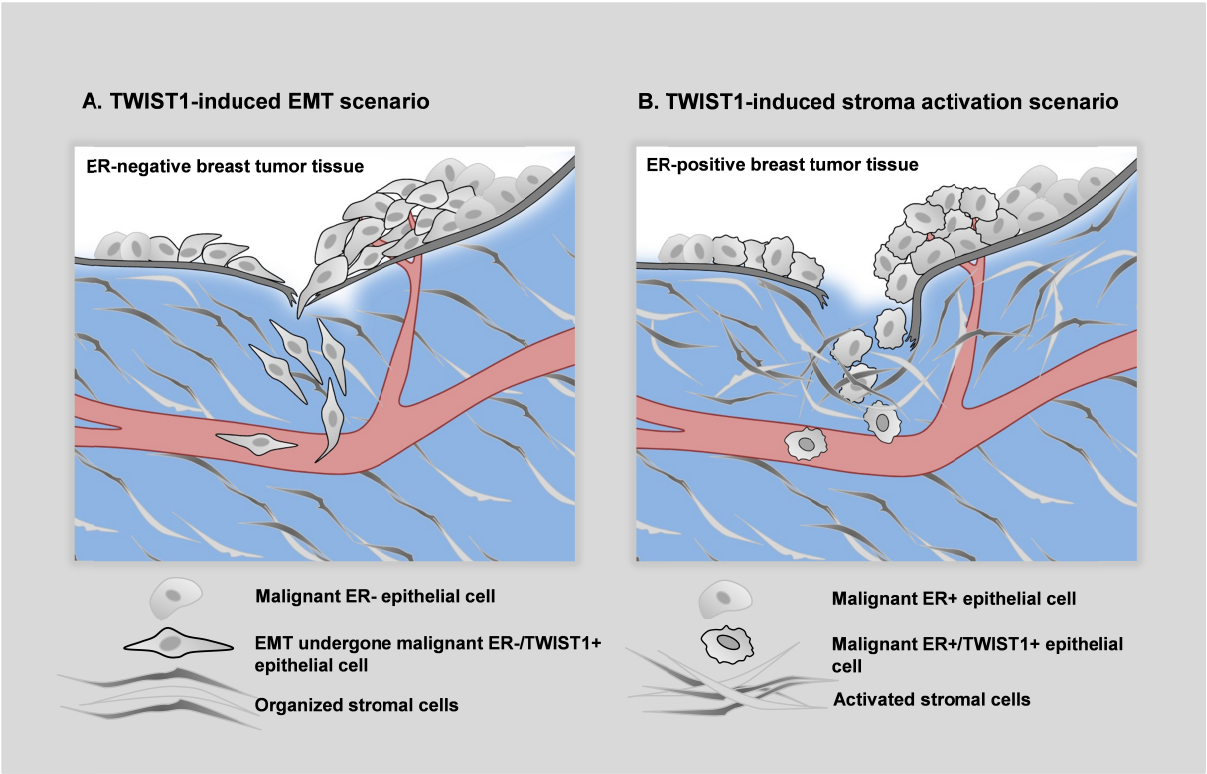


Figure 9.3 Schematic representation of two proposed scenarios of TWIST-induced disease progression in breast cancer tissue.

In this scenario, TWIST1 expression causes the malignant epithelial cells to secrete *stroma activating factors* or tumor cells modulate TWIST1 expression in the stroma leading to activated stroma in tumor tissue. Such factors allow the stroma of tumor tissue to become loose-fitting and to secrete

metalloproteinase enzymes, which not only degrade ECM of the tissue but also breach the basement membrane thereby facilitating malignant epithelial cells to escape into the blood circulation without undergoing full-blown EMT-like changes. The clues about how malignant epithelial cells themselves or induce stromal cells to secrete so called stroma activating factors come from the work of Eckert and colleagues. They showed that TWIST1 expression induces the development of subcellular protrusions known as invadopodia[45]. These invadopodia-bearing cells secrete metalloproteinases[45], which not only degrade ECM of the tissue but also activate stromal components in the tissue. It is therefore clear that more work is needed to completely unfold these scenarios in the context of breast cancer.

A fundamental question with regard to the two proposed TWIST1-related scenarios of disease progression would be: what factor would determine which TWIST1 scenario should prevail in the context of breast cancer progression? Based on our data, we think that ER would be a determining factor. In ER-positive breast tissue TWIST1-mediated disease progression would therefore follow, the *TWIST1-induced stroma activation scenario'*, whereas ER-negative breast tissue favors the *'TWIST1-induced EMT scenario'*. To provide functional evidence that '*TWIST1-induced stroma activation'* contributes to ER-positive/luminal breast cancer progression, TWIST1 expression has to be manipulated to determine if and how it can affect the tumor stroma cells to cooperate with epithelial cells in the process of invasion and metastasis. Once evidence is provided that the '*TWIST1-induced stroma activation scenario'* is of importance, drugs targeting this route may prevent TWIST1-mediated metastasis in luminal breast cancer.

Human breast cancer cell lines as models to understand breast cancer

In **chapter 7** and **chapter 8** of this thesis, we performed molecular characterization of human breast cancer cell lines to evaluate *in vitro* and *in vivo* subtype-specific biological differences and to further establish how reliably the *in vitro* cultured human breast cancer cell lines recapitulate the biology of the primary breast tumors. In **chapter 7**, we extended the molecular characterization of our large panel of breast cell lines to study differential expression of 725 human miRNAs using a microarray carrying 1,344 LNA™ modified oligonucleotide capture probes. We studied the association of these miRNAs with clinically important factors such as the intrinsic molecular subtypes and ER, PGR, as well as ERBB2 expression and with genetic factors such as *TP53*, *PIK3CA*, and *E-cadherin* mutation status as well as global genomics data such as SNP copy number variations and global gene expression.

To identify molecular subtypes we first performed a microarray gene expression analysis to retrieve global gene expression patterns of our collection of 51 human breast cancer cell lines. Pearson correlation as well as cluster analysis using the intrinsic gene-set defined by Perou and colleagues[46] revealed three prominent independent clusters of cell lines. These clusters mainly represent luminal-subtype, basal-like, and normal-like/claudin-low subtypes. Cell lines with an ERBB2 overexpression representing the ERBB2-positive intrinsic subtype were also observed. However, these cell lines did not constitute an independent group and mostly clustered together with the luminal-type cell lines. Based unsupervised analysis of 413 miRNAs we observed a similar division

into two main clusters driven by ER status. MiRNAs that were differentially expressed between ER-positive and ER-negative cell lines showed a considerable overlap with those differentially expressed between primary tumors of different ER status[47]. However, not all miRNAs are concordantly differentially expressed. Notably, expression of *hsa-miR-100* miRNA was high in ER-positive primary breast tumors[48] while in the cell lines this miRNA displayed high expression in ER-negative cell lines. Such discrepancies could be explained by a significant contribution of the tumor microenvironment to the expression of some individual miRNAs[47]. The second miRNA signature driving the unsupervised division of breast cancer cell lines involved miRNAs that were previously implicated in the EMT process (*hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, hsa-miR-141*) in metastatic breast tumors and are associated with the normal-like/claudin-low cell lines, which also show a prominent EMT and stem cell-like phenotype[49,50].

To evaluate relevant biological parameters of tumor progression (E.g., tumor growth, incidence and type of metastasis) in the final chapter **(chapter 9)** of this thesis, we developed xenograft mice models from 21 breast cancer cell lines representing four molecular subtypes. Each cell line was first equipped with a luciferase reporter gene to image and to follow breast cancer cells in growing xenografts in living animals in real-time manner. The ability of most of these cell lines to grow as xenografts suggests that the cell lines, despite of their *in vitro* growth for many years, still retain their natural capacity to exploit and communicate with the host microenvironment. Consistent with clinical breast tumors, the growth rate of the xenografts from normal-like/claudin-low cell lines was high and mice bearing these xenografts showed poor tumor-related survival. In contrast to observation made in the clinic, ERBB2-positive cell lines behave significantly less aggressive compared to cell lines of other molecular subtypes. Furthermore, in contrast to the clinical observation that molecular subtypes of primary breast tumors exhibits preferences to relapse to specific sites (i.e. bone, lungs, brain)[51], breast cancer cell lines in xenografts did not show such subtype-specific relapse to specific organs. Rather, cell lines from all molecular subtypes shared common organs for metastasis in the mouse, which mostly were lymph nodes and lungs. One obvious explanation is that the small number of cell lines that we were able to use to generate xenografts, which have not given us the power to study this property. Importantly, xenografts from these cell lines grow in recipient mice in a reasonable time span, which makes these cell line xenografts unique models to screen potential breast cancer drugs in the presence of host stroma.

As a whole, based on the analysis by others and us, it is concluded that breast cancer cell lines despite being propagated *in vitro* for decades, to a significant extent, reflect the features of primary breast tumor cells *in vivo* E.g., with regard to genomics, Neve and colleagues in 2006 using a BAC array CGH-based comparative analysis[52] reported recurrent gains at chromosomes q1, 8q, 11q, 12q, 17q, and 20q and losses at chromosomes 3q, 13q, 16q, 17p, and 22q[52,53] in both 185 primary breast tumors and 51 human breast cancer cell lines. Using SNP copy number in a global analysis, we also observed strikingly similar patterns of chromosomal aberrations in 303 primary tumors[54] and 56 breast cancer cell lines (unpublished). Furthermore, global mRNA and miRNA expression patterns divide cell lines and primary tumors into two major groups dictated by ER status. These immense ER-mediated differences have been recognized many years ago and confirmed that each of these

molecular subtypes represents a unique biology with a different prognosis and treatment options[55]. With regard to aggressive behavior of cell lines grown in mice, in line with observations found in primary clinical breast cancer, basal-like and in particular normal-like/claudin-low cell lines behave more aggressively. The latter is also in line with the requirement of EMT for rapid dissemination of ER-negative tumors. Overall, this indicates that the cell lines do not substantially accumulate new recurrent genomic and phenotypic alterations due to *in vitro* cultivation, but rather they globally have a similar genomic landscape as primary tumors from which they are derived.

Despite of a general concordance between breast cancer cell lines and primary breast cancer tumors, this cell lines collection with regard to certain features represents a biased population. For example, ER-negative and triple-negative cell lines are over represented in the collection of cell lines while, in primary breast cancers, a minority of tumors (10-25%) is ER-negative or triple-negative[56]. Another notable difference between cell lines and primary tumors is the absence of a luminal A and luminal B subdivision in the cell lines, whereas these two molecular subtypes have repeatedly been observed in primary breast tumors even from various ethnic backgrounds[57]. Moreover, tumors of the luminal A and B subtype a display diverse and distinct prognosis[57]. Furthermore, a substantial numbers of cancer cell lines exhibits a mesenchymal-cell like morphology, which according to pathologists is a rare phenomenon to be seen during histological examination of primary breast tumors. Additionally, the mutational frequency of certain genes is significantly higher in cell lines compared with primary breast tumors. For instance, mutation frequency of the *TP53* gene is approximately twice as high in cell lines as compared with primary breast tumors[58]. Similar to *TP53* mutations, *E-cadherin* mutations and amplification of *ERBB2* are more frequently observed in breast cancer cell lines compared to primary tumors. Also with regard to aggressiveness, ERBB2-overexpressing cell lines do not behave as anticipated from ERBB2-positive clinical specimens and also organ-specific relapse in cell lines is not molecular subtype-related as it has been observed for clinical specimen. The observed differences between cell lines cultured *in vitro* or grown as xenografts might reflect a general aggressive nature of more dedifferentiated tumors, which might more easily adapt to *in vitro* conditions. Notably, the cell lines grown *in vitro* or as in pure tumor xenografts lack a stromal component and therefore obvious stromal-derived genes or miRNAs cannot be detected and thus potential stromal-related biology cannot be studied in cell lines. Although these cell lines could be very useful, one must take into account all these pros and corns of the breast cancer cell lines before considering them to model clinical breast cancer.

REFERENCES

1. Birch, J.M. et al. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* **20**, 4621-8 (2001).
2. Varley, J.M. et al. Characterization of germline TP53 splicing mutations and their genetic and functional analysis. *Oncogene* **20**, 2647-54 (2001).

3. Hall, J.M. et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-9 (1990).
4. Neuhausen, S. et al. Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. *Nat Genet* **13**, 126-8 (1996).
5. Robson, M.E. & Offit, K. New BRCA2 mutation in an Ashkenazi Jewish family with breast and ovarian cancer. *Lancet* **350**, 117-8 (1997).
6. Lynch, E.D. et al. Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* **61**, 1254-60 (1997).
7. Jenne, D.E. et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* **18**, 38-43 (1998).
8. Berx, G., Becker, K.F., Hofler, H. & van Roy, F. Mutations of the human E-cadherin (CDH1) gene. *Hum Mutat* **12**, 226-37 (1998).
9. Stratton, M.R. & Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**, 17-22 (2008).
10. Meijers-Heijboer, H. et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
11. Rahman, N. et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**, 165-7 (2007).
12. Renwick, A. et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-5 (2006).
13. Seal, S. et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-41 (2006).
14. Varon, R. et al. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467-76 (1998).
15. Hollestelle, A., Wasielewski, M., Martens, J.W. & Schutte, M. Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* **20**, 268-76 (2010).
16. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
17. Ghoussaini, M. et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-8 (2012).
18. Park, J.H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42**, 570-5 (2010).
19. Vahteristo, P. et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**, 432-8 (2002).
20. Weischer, M., Bojesen, S.E., Ellervik, C., Tybjaerg-Hansen, A. & Nordestgaard, B.G. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol* **26**, 542-8 (2008).
21. Meijers-Heijboer, H. et al. The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet* **72**, 1308-14 (2003).

22. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* **74**, 1175-82 (2004).
23. Wasielewski, M. CHEK2 1100delC and polygenic susceptibility to breast cance and colorectal cancer. *PhD theses: Departmement of Medical Oncology, ErasmusMC.* (2009).
24. Mitchell, R.J., Farrington, S.M., Dunlop, M.G. & Campbell, H. Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. *Am J Epidemiol* **156**, 885-902 (2002).
25. Kolodner, R.D. et al. Germ-line msh6 mutations in colorectal cancer families. *Cancer Res* **59**, 5068-74 (1999).
26. Zhao, Y.S. et al. Meta-analysis of MSH6 gene mutation frequency in colorectal and endometrial cancers. *J Toxicol Environ Health A* **72**, 690-7 (2009).
27. Peltomaki, P. Lynch syndrome genes. *Fam Cancer* **4**, 227-32 (2005).
28. Wasielewski, M. et al. CHEK2 1100delC is a susceptibility allele for HNPCC-related colorectal cancer. *Clin Cancer Res* **14**, 4989-94 (2008).
29. Bernstein, B.E. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
30. Rouleau, M., Patel, A., Hendzel, M.J., Kaufmann, S.H. & Poirier, G.G. PARP inhibition: PARP1 and beyond. *Nat Rev Cancer* **10**, 293-301 (2010).
31. Smid, M. et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol* **24**, 2261-7 (2006).
32. Cowper-Sal Lari, R. et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* (2012).
33. Meyer, K.B. et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* **6**, e108 (2008).
34. Ahmadiyeh, N. et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A* **107**, 9742-6 (2010).
35. Freedman, M.L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* **43**, 513-8 (2011).
36. Fasching, P.A. et al. The role of genetic breast cancer susceptibility variants as prognostic factors. *Hum Mol Genet* **21**, 3926-39 (2012).
37. Soini, Y. et al. Transcription factors zeb1, twist and snai1 in breast carcinoma. *BMC Cancer* **11**, 73 (2011).
38. van Nes, J.G. et al. Co-expression of SNAIL and TWIST determines prognosis in estrogen receptor-positive early breast cancer patients. *Breast Cancer Res Treat* **133**, 49-59 (2012).
39. Roman-Perez, E. et al. Gene expression in extratumoral microenvironment predicts clinical outcome in breast cancer patients. *Breast Cancer Res* **14**, R51 (2012).
40. Yang, J. et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-39 (2004).
41. Yang, M.H. et al. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat Cell Biol* **12**, 982-92 (2010).

42. Pozharskaya, V. et al. Twist: a regulator of epithelial-mesenchymal transition in lung fibrosis. *PLoS One* **4**, e7559 (2009).
43. Sarrio, D. et al. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res* **68**, 989-97 (2008).
44. Blick, T. et al. Epithelial mesenchymal transition traits in human breast cancer cell lines. *Clin Exp Metastasis* **25**, 629-42 (2008).
45. Eckert, M.A. et al. Twist1-induced invadopodia formation promotes tumor metastasis. *Cancer Cell* **19**, 372-86 (2011).
46. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747-52 (2000).
47. Foekens, J.A. et al. Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proc Natl Acad Sci U S A* **105**, 13021-6 (2008).
48. Mattie, M.D. et al. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer* **5**, 24 (2006).
49. Gregory, P.A. et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593-601 (2008).
50. Korpal, M., Lee, E.S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**, 14910-4 (2008).
51. Smid, M. et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**, 3108-14 (2008).
52. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-27 (2006).
53. Chin, K. et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**, 529-41 (2006).
54. Smid, M. et al. Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. *Breast Cancer Res Treat* **128**, 23-30 (2011).
55. Brenton, J.D., Carey, L.A., Ahmed, A.A. & Caldas, C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* **23**, 7350-60 (2005).
56. Jonat, W. et al. Effectiveness of switching from adjuvant tamoxifen to anastrozole in postmenopausal women with hormone-sensitive early-stage breast cancer: a meta-analysis. *Lancet Oncol* **7**, 991-6 (2006).
57. Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**, 8418-23 (2003).
58. Hollestelle, A. et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res Treat 121, 53-64 (2010).

Chapter 10

Summary

Samenvatting

خلاصه

SUMMARY

Breast cancer is the second leading cause of cancer-related deaths in the western world and one of the most common diseases worldwide. About 1.4 million new cases of invasive breast cancer are expected to be diagnosed and approximately 0.7 million women will die from this disease in 2012. Breast cancer is a genetic disease, the onset of which has been associated with mutations in functionally important genes commonly known as breast cancer susceptibility genes. These genes can either be inherited or acquired somatically in the tumor tissue. Based on associated breast cancer risk and population prevalence, the breast cancer susceptibility genes can be classified as: (1) high-risk breast cancer susceptibility genes that are very rare in the general population but associated with high risk; (2) moderate-risk breast cancer susceptibility genes that are also rare and associated with moderate risk; and (3) low-risk breast cancer susceptibility loci, which are common in the general population but associated with a small increase in risk. Biological disease mechanisms that underlie the risk conferred by the high-risk and moderate-risk susceptibility genes and their role in the progression of the disease have previously been well studied. But, how low-risk breast cancer loci confer their increased breast cancer risk in patients is poorly understood. Moreover, whether these loci can predict the course of the disease in primary breast cancer patients has also not been previously addressed. The complete clinical significance and the biological function of the newly discovered low-risk breast cancer susceptibility loci is still to be understood. Furthermore, a wide variety of well-characterized *in vivo* models to study genetic and biological factors are lacking, hampering understanding of their role in disease progression.

In the **first part** of this thesis, we applied a candidate gene approach to discover additional genetic aberrations in familial breast cancer and to gain more insight into the biology and prognosis of known low-risk loci and candidate genes. In **chapter 3** of this thesis, we aimed to discover whether mutations in the *MSH6* gene were associated with breast cancer risk in breast and colorectal cancer (HBCC) families. The HBCC phenotype was of interest since one-fifth of HBCC families carry a *CHEK2* 1100delC mutation and this mutation was suggested to confer hereditary breast cancer risk in a polygenic setting. We therefore hypothesized that mutations in the *MSH6* gene might act as partners of the *CHEK2* 1100delC mutation. The rationale to perform this study was based on the fact that both of these genes have previously been linked to the Lynch Syndrome. Thus, we analyzed the entire *MSH6* coding sequence for genetic alterations in blood-derived DNA of 68 HBCC breast cancer families. All rare *MSH6* variants, with population frequencies below 1%, were identified in 11.8% of HBCC breast cancer families whereas the same variants were identified in only 1.5% of population controls. This suggests that rare *MSH6* variants are associated with HBCC breast cancer ($P \leq 0.00001$). However, screening of the entire *MSH6* coding sequence in 68 non-HBCC breast cancer families showed a similar association (8.8% vs. ~1.4% in controls, $P \leq 0.001$), suggesting that rare *MSH6* variants are not confined to HBCC breast cancer. Together, our data suggest that the rare *MSH6* variants may predispose to familial breast cancer. However, none of the rare *MSH6* variants are obviously pathogenic, suggesting that a more subtle disease mechanism may operate in breast carcinogenesis.

Then we focused on (i) how low-risk breast cancer susceptibility loci confer their increased risk of breast cancer and (ii) whether low-risk breast cancer susceptibility loci can affect the natural course of the disease progression in breast cancer patients. We genotyped nine candidate SNPs capturing seven risk-associated loci (FGFR2-rs2981282, LSP1-rs3817198, MAP3K1-rs889312, TOX3-rs3803662, H19-rs2107425, 2q35-rs13387042, and 8q24-rs13281615) in 40 human breast cancer cell lines **(chapter 4)** and in a retrospective cohort of 2,480 primary breast cancers with complete clinical follow-up **(chapter 5)**. To explore underlying disease mechanisms, we tested the most plausible hypothesis in which low-risk breast cancer susceptibility loci were suspected to confer their increased risk, which is via the modulation of expression of the nearest genes. The minor allele frequency (MAF) of these loci was indeed higher in the breast cancer cell lines and in the tumors compared with blood-derived DNA of a control population of European ancestry. Our results revealed that except for SNP rs3803662, which was associated with lower expression of the *TOX3* gene in the clinical cohort, all tested low-risk breast cancer susceptibility loci generally do not affect the expression levels of the nearest genes in the cell lines as well as in the tumor samples. Pertaining to prognosis of these loci, we found that the disease course in patients is also largely not associated with low-risk breast cancer loci except for *H19* SNP (rs2107425; G/A) near the *H19* gene. The homozygous risk alleles (AA) of *H19* SNP were associated with shorter metastasis-free survival. Thus, while these low-risk breast cancer susceptibility loci have been repeatedly associated with increased risk of breast cancer we conclude that the low-risk breast cancer susceptibility loci in general do not strongly associate with altered expression of the nearest gene nor are they linked to altered disease progression **(chapter 5)**.

In the **second part** of the thesis, we investigated breast cancer subtype-specific biological factors such as genes and micro-RNAs using either primary breast tumors or *in vitro* propagated breast cancer cell lines. Moreover, we developed breast cancer cell line xenograft mice models to determine whether subtype-related clinical behavior is maintained in these models. And also whether these models can be used to explore the biology of the clinical factors and to investigate interference of currently existing as well as novel targeted drugs. Using patient's primary tumor samples, we investigated the prognostic significance of *TWIST1*, a well-known transcription factor and marker of epithelial to mesenchymal transition (EMT), in a cohort of 1,427 patients with extensive clinical follow-up. In addition, using available gene expression array data we explored the underlying biology. Our study revealed that *TWIST1* mRNA expression is an independent prognostic marker for poor prognosis in ER-positive lymph-node negative breast cancer. Additionally, our study provided evidence that the adverse effect of TWIST1 is likely connected to the tumor microenvironment and/or high stromal composition of tumor tissue **(chapter 6)**.

To show whether breast cancer cell lines as a model can mimic clinical tumors at mRNA and miRNA expression levels, we performed global mRNA and miRNA expression analyses of 51 human breast cancer cell lines **(chapter 7)**. Both the mRNA and the miRNA expression profiles of the cell lines could discriminate ER-positive and ER-negative breast cancer cell lines, which is in line with the commonly accepted hypothesis that the ER-positive and ER-negative tumors are biologically very distinct. Additionally, we revealed that certain miRNAs were differentially expressed among

intrinsic breast cancer subtypes, among EMT status of the cell lines and also between *wild-type* cell lines versus those carrying mutations in tumor suppressor and oncogenes commonly present in clinical breast tumors. These results further underscore that breast cancer cell lines retain biological features that have been described for clinical tumors. Therefore, we argue that the breast cancer cell line model is of relevance for studying breast cancer biology. Moreover, our miRNA dataset generated from the cell lines is unique, and can be used to elucidate the underlying biology of clinically important miRNAs in breast cancer.

Finally, we evaluated the propensity of our breast cancer cell line model to grow and metastasize in the mouse microenvironment and to assess its use as a preclinical xenograft model to study whether their metastatic behavior resembles the behavior of breast tumors in patients. For this, we stably transduced from each of the four common intrinsic molecular subtypes at least four breast cancer cell lines with a luciferase reporter gene and implanted them orthotopically in nude mice. Our findings revealed that in our hands the majority of the cell lines were able to grow and gave rise to xenograft tumors in mice (tumor take rate = 87.5%). As observed in clinical tumors, normal-like/claudin-low cell lines displaying features of EMT and adult mammary stem cells showed a more aggressive growth pattern in mice compared to other subtypes of cells (Mann-Whitney test; *P*-value =0.001). Following metastasis, we found that lymph-nodes, lungs, and liver were the most frequently invaded organs by all types of cell lines irrespective of their intrinsic subtype **(chapter 8)**. From these data we conclude that despite of being propagated *in vitro* for years, the human breast cancer cell lines retain their inherent potential to grow and metastasize in the mouse tissue microenvironment. We therefore argue that a collection of human breast cancer cell lines is a relevant model, which can not only be applied as an initial discovery tool to gain more global genetics and biological insights into human breast carcinogenesis, but can also be employed as a preclinical breast cancer xenograft model to screen potential breast cancer drugs.

SAMENVATTING

Borstkanker is één van de meest voorkomende ziektes in de wereld. In de westerse wereld is het op één na de belangrijkste kankergerelateerde doodsoorzaak. In 2012 zal bij 1.4 miljoen vrouwen de diagnose invasief borstkanker gesteld worden. Ongeveer de helft van deze vrouwen zal op termijn aan deze ziekte overlijden. Borstkanker is een genetische ziekte. Mutaties in functioneel belangrijke genen die vrouwen ontvankelijk maken voor het krijgen van borstkanker, de borstkanker predispositiegenen, spelen daarin een belangrijke rol. De predispositiegenen krijg je via overerving of ze kunnen ontstaan tijdens de ontwikkeling van de tumor. Borstkanker predispositiegenen kunnen op basis van het geassocieerde borstkankerrisico en hun voorkomen in de algehele populatie geclassificeerd worden als: (1) hoog-risico borstkanker predispositiegenen; zeldzaam in de algehele populatie, maar ze vertonen een sterke associatie met het optreden van borstkanker; (2) gematigd-risico borstkanker predispositiegenen, deze genen zijn minder zeldzaam en worden geassocieerd met een intermediair risico op het optreden van borstkanker; en (3) laag-risico borstkanker predispositiegenen; komen vaak voor in de algehele populatie en zijn geassocieerd met slechts een licht verhoogd risico op het optreden van borstkanker.

De biologische mechanismen die ten grondslag liggen aan het risico van de hoog-risicogenen, en in mindere mate de gematigd-risicogenen, en hun rol in de progressie van de ziekte zijn in het verleden uitgebreid bestudeerd. Het biologisch mechanisme van de laag-risico borstkankergenen bij het ontstaan van borstkanker is nog onduidelijk. Ook is nog niet eerder vastgesteld of deze laag-risico borstkankergenen het verloop van de ziekte na de diagnose in borstkankerpatiënten kan voorspellen. Helaas ontbreekt het aan goed gekarakteriseerde *in vivo* modelsystemen om genetische en biologische factoren te bestuderen. Dit belemmert het onderzoek naar hun rol in ziekteprogressie.

In het **eerste deel** van het proefschrift hebben we een specifieke benaderingswijze gebruikt om genetische afwijkingen in familiair borstkanker te ontdekken en om inzicht te krijgen in de biologie van bekende laag-risicogenen en potentiële kandidaatgenen en hun effect op prognose. In **hoofdstuk 3** van dit proefschrift was het doel om een mogelijke relatie tussen mutaties in het *MSH6* gen met het risico tot het optreden van borstkanker in reeds geïdentificeerde borst- en darmkanker (HBCC) families vast te stellen. Het HBCC phenotype is interessant omdat één vijfde van de HBCC families drager is van een *CHEK2* 1100delC mutatie. Deze mutatie verhoogt mogelijk in combinatie met een mutatie in een of meer andere genen het risico voor het optreden van borstkanker. Onze hypothese was daarom dat mutaties in het *MSH6* gen het effect van de *CHEK2* 1100delC mutatie bij de ontwikkeling van borstkanker in HBCC families kan versterken. De reden voor deze studie was het feit dat beide genen al eerder in verband waren gebracht met het Lynch syndroom. Wij hebben de gehele coderende sequentie van het *MSH6* gen geanalyseerd op genetische veranderingen in 68 HBCC families. Zeldzame (<1%) *MSH6* varianten werden geïdentificeerd in 11.8% van de HBCC borstkankerfamilies, terwijl deze zelfde varianten in slechts 1.5% van de controle populatie werden gevonden. Dit suggereert dat zeldzame *MSH6* varianten geassocieerd zijn met een risico op HBCC borstkanker ($P \leq 0.00001$). Echter, screening van de gehele coderende sequentie van het *MSH6* gen in 68 non-HBCC borstkankerfamilies liet een

vergelijkbare associatie zien (8,8% t.o.v. ~1,4% in de controlegroep, $P \leq 0{,}001$). Dit resultaat suggereert dat zeldzame *MSH6* varianten niet beperkt zijn tot HBCC borstkanker. Samenvattend suggereren onze data dat zeldzame *MSH6* varianten in het algemeen vrouwen ontvankelijk maken voor familiair borstkanker. Geen van de zeldzame *MSH6* varianten is echter overduidelijk ziekteverwekkend, mogelijk ligt er een gecompliceerd ziektemechanisme aan ten grondslag tijdens het ontstaan van borstkanker door dergelijke varianten.

Vervolgens hebben we onze aandacht gericht op de vragen (i) hoe laag-risico borstkanker predispositiegenen het risico op borstkanker verhogen (ii) of laag-risico borstkanker predispositiegenen de natuurlijke progressie van de ziekte in borstkankerpatiënten kunnen beïnvloeden. Samengevat hebben wij onderzocht of dergelijke single-nucleotide polymorfismes (SNPs) de prognose van een patiënt kunnen voorspellen nadat de ziekte geconstateerd is. Negen kandidaat SNPs in of in de buurt van een gen op zeven met risico geassocieerde loci (FGFR2-rs2981282, LSP1-rs3817198, MAP3K1-rs889312, TOX3-rs3803662, H19-rs2107425, 2q35-rs13387042, and 8q24-rs13281615) werden daarom gegenotypeerd in 40 humane borstkankercellijnen **(hoofdstuk 4)** en in een retrospectief cohort van 2480 primaire borsttumoren van patiënten van wie de volledige klinische follow-up bekend is **(hoofdstuk 5)**. Om de onderliggende ziektemechanismen te onderzoeken hebben we de meest waarschijnlijke hypothese getest, waarbij laag-risico borstkanker SNPs een hoger risico opleveren door een verandering van de expressie van het gen dat het dichtst in de buurt van de SNP ligt. De frequentie van het minst voorkomende allel voor deze loci was inderdaad hoger in de tumorcellijnen en de tumoren dan in het bloed van de controle populatie van Europese afkomst. Dit bevestigt dat ook in ons cohort de allelen waarop deze SNPs gelegen zijn een verhoogd risico op borstkanker geven. Onze resultaten toonden verder aan dat alle zeven onderzochte laag-risico borstkanker predispositieloci over het algemeen niet het expressieniveau van het dichtstbijzijnde gen beïnvloeden. Dat gold voor de cellijnen en de tumorweefsels. Een uitzondering hierop was SNP rs3803662, welke geassocieerd was met een lagere expressie van het *TOX3* gen. Met betrekking tot de prognose van deze loci vonden wij dat het ziekteverloop in patiënten niet is geassocieerd met laag-risico borstkankerloci, behalve in het geval van de *H19* SNP (rs2107425 G/A) dat gelegen is bij het *H19* gen. Homozygotie van het risico-allel (genotype AA) van de *H19* SNP was geassocieerd met een kortere metastase-vrije overleving. Dus, hoewel deze laag-risico borstkanker predispositieloci herhaaldelijk geassocieerd zijn met een verhoogd risico op borstkanker moeten we concluderen dat deze SNPs geen effect hebben op het natuurlijk beloop van de progressie van borstkanker. Ook vonden we geen aanwijzingen dat deze SNPs over het algemeen een sterke associatie hebben met een gewijzigde genexpressie van het dichtstbijgelegen gen, de tot nog toe meest gangbare hypothese **(hoofdstuk 5)**.

In het **tweede deel** van het proefschrift hebben wij borstkankersubtype-specifieke biologische factoren, zoals genen en microRNAs, onderzocht in primaire borsttumoren en gekweekte borstkankercellijnen. Als laatste hebben wij muizen xenotransplantatie modellen ontwikkeld van borstkankercellijnen. Dit om te bepalen of het klinisch gedrag van de borstkanker-subtypes in de

muis behouden blijft en om in de toekomst functioneel onderzoek te kunnen doen aan verschillende borstkankersubtypes.

In klinische borsttumorweefsels hebben wij de prognostische betekenis van *TWIST1*, een bekende transcriptiefactor en marker voor epitheliale naar mesenchymale transitie (EMT), onderzocht in een cohort van 1427 primaire borstkankerpatiënten met uitgebreide klinische follow-up. Daarnaast hebben wij de onderliggende biologie bestudeerd met behulp van beschikbare genexpressie array data. Onze studie wees uit dat *TWIST1* mRNA expressie een onafhankelijke prognostische marker was voor een slechtere prognose in ER-positief, lymfeklier negatief borstkanker. Bovendien liet onze studie zien dat het ongunstige effect van TWIST1 waarschijnlijk samenhangt met de micro-omgeving van de tumor en/of een rijke stromale samenstelling van het tumorweefsel **(hoofdstuk 6)**.

Om te onderzoeken of een model van borstkankercellijnen de mRNA en miRNA expressieniveaus van klinische tumoren kan imiteren hebben wij genoom-brede mRNA and miRNA expressieanalyses uitgevoerd voor 51 humane borstkankercellijnen **(hoofdstuk 7)**. Zowel de mRNA als de miRNA expressieprofielen van de cellijnen konden de ER-positieve van de ER-negatieve borstkankercellijnen onderscheiden, dat is in overeenstemming met de algemeen geaccepteerde hypothese dat ER-positieve en ER-negatieve tumoren biologisch verschillend zijn. Daarnaast hebben wij kunnen aantonen dat bepaalde miRNAs differentieel geëxpresseerd werden tussen de verschillende moleculaire subtypes van borstkanker en dat ze overeen kwamen met de EMT status van de cellijnen. Ook lieten specifieke miRNAs verschillende expressieniveaus zien in *wild-type* cellijnen vergeleken met cellijnen die mutaties dragen in frequent gemuteerde tumorsuppressorgenen en oncogenen. Deze resultaten bevestigen dat borstkankercellijnen de biologische kenmerken behouden die beschreven worden voor primaire tumoren uit patiënten. Wij stellen daarom dat het borstkankercellijn model relevant is om de biologie van borstkanker te bestuderen. De miRNA dataset die we gegenereerd hebben voor de cellijnen is uniek en kan gebruikt worden om de onderliggende biologie van klinisch belangrijke miRNAs in borstkanker te ontrafelen.

Als laatste hebben we in de intrinsieke capaciteit van de borstkankercellijnen om in een micro-omgeving te groeien en uit te zaaien onderzocht met een muis-model. We hebben de toepasbaarheid van ons preklinisch xenotransplantatie model geëvalueerd door het metastaseringsgedrag in de muis te vergelijken met dat in de patiënt. Voor dit doel hebben we minimaal vier borstkankercellijnen van elk van de vier intrinsieke moleculaire subtypes stabiel getransduceerd met een luciferase reportergen en vervolgens geïmplanteerd in het orgaan van origine in naakte muizen. Onze bevindingen toonden aan dat de meerderheid van de gebruikte cellijnen (87%) in staat waren om te groeien en tumoren te vormen in de muis. Zoals ook waargenomen in klinische tumoren, hadden normal-like/claudine-lage cellijnen, welke kenmerken van EMT en borststamcellen vertonen, een agressiever groeipatroon in muizen dan de cellijnen van de andere borstkankersubtypes. Wij vonden dat de lymfeklieren, longen en lever het meest frequent aangedaan waren, ongeacht het intrinsieke borstkankersubtype **(hoofdstuk 8)**. Uit deze data concluderen wij dat humane borstkankercellijnen, ondanks het feit dat ze vele jaren in kweek

zijn, hun inherente mogelijkheid om te groeien en te metastaseren behouden in de micro-omgeving in de muis. Wij bepleiten daarom dat een collectie van humane borstkankercellijnen een relevant model is. De toepassing van het borstkanker xenotransplantatie muismodel hoeft zich niet te beperken tot het verkrijgen van genetische en biologische inzichten in het ontstaan van borstkanker in de mens. Het kan ook preklinisch gebruikt worden om de effectiviteit van potentiële borstkankermedicijnen te screenen.

مقالہ کا عنوان: **انسانی چھاتی کا سرطان اور اس کے جنیاتی، حیاتیاتی اور مرض میں پھیلاؤ کے تشخیصی عوامل**

خلاصہ

چھاتی کا سرطان مغربی دنیا میں سرطان سے منسلک اموات کی دوسری اہم وجہ ہے۔ایک اندازے کیمطابق 2012ء میں تقریباً 14لاکھ چھاتی کے سرطان کے کینسر کی تشخیص ہوگی اور تقریباً 7لاکھ خواتین کی موت متوقع ہے۔ چھاتی کا سرطان ایک جینیاتی مرض ہے جو کہ اہم جینز میں تبدیلی رونما ہونے کی وجہ سے ہوتا ہے۔ان تبدیل ہونے والے جینز کو چھاتی کے سرطان کا سبب بننے والے جینز کہا جاتا ہے۔ یہ جینز یا تو وراثتی طور پر منتقل ہوتے ہیں یا ان میں تبدیلی سے صرف رسولی والے ٹشو میں ہی رونما ہوتی ہے۔ان جینز کو ان سے لاحق ہونے والے سرطان کے خطرے اور ان کی آبادی میں موجودگی کی شرح کی بنیاد پر مندرجہ ذیل تین گروہوں میں تقسیم کیا گیا ہے۔

1۔ چھاتی کے سرطان کے انتہائی خطرے والے جینز! یہ جینز بھی عام آبادی میں کافی نایاب ہیں۔مگر ان سے سرطان کے رونما ہونے کا خطرہ کافی زیادہ ہوتا ہے۔

2۔ چھاتی کے سرطان کے درمیانے درجے والے جینز! یہ جینز بھی عام آبادی میں نایاب ہیں مگر ان سے سرطان رونما ہونے کا خطرہ درمیانے درجے کا ہوتا ہے۔

3۔ چھاتی کے سرطان کے کم خطرے والے جنز یا لوسائی! یہ جینز یا لوسائی عام آبادی میں کثیر تعداد میں پائے جاتے ہیں مگر ان سے سرطان رونما ہونے کا خطرہ بہت کم ہوتا ہے۔

اس سے پہلے چھاتی کے سرطان کے انتہائی اور درمیانے درجے کے خطرے والے جینز کے حیاتیاتی طریقہ کار اور ان کے سرطان میں پھیلاؤ کے کردار پر کافی مفصل تحقیق ہو چکی ہے۔لیکن چھاتی کے کم خطرے والے لوسائی مریض میں چھاتی کے سرطان کا خطرہ کیسے بڑھاتے ہیں اور ابتدائی سرطان کے پھیلاؤ کی رفتار پر کیسے اثر انداز ہوتے ہیں بہت کم تحقیق ہوئی ہے اس کے علاوہ ان نئے دریافت شدہ چھاتی کے سرطان کے کم خطرے والے لوسائی کی طبعی خصوصیات اور ان کے حیاتیاتی طریقہ کار کے متعلق بھی بہت کم معلومات ملتی ہیں۔مزید یہ کہ ان کے حیاتیاتی اور جنیاتی عوامل کے کردار پر تحقیق کرنے کے لئے موضوں اور مصدقہ ماڈلز بھی دستیاب نہیں ہیں جس کی وجہ سے چھاتی کے سرطان کے پھیلاؤ کی رفتار کا تعین کرنا مشکل ہو رہا ہے۔

اس مقالے کے پہلے حصہ میں ہم نے ایک امیدوار جین کا نقطہ نظر اپناتے ہوئے چھاتی کے موروثی سرطان کے مریضوں کے جنیاتی تبدیلیاں تلاش کی ہیں۔اس کے ساتھ ساتھ ہم نے چھاتی کے سرطان کے کم خطرے والے لوسائی اور امیدوار جینز کے حیاتیاتی طریقہ کار اور ان کے بیماری کے پھیلاؤ کی رفتار اور تعین کرنے کے کردار پر تحقیق کی ہے۔مقالے کے باب نمبر 3 میں ہم نے ایم۔ایس۔ایچ 6(*MSH6*) جین میں تبدیلیوں سے منسلک سرطان پر چھاتی اور انتڑیوں کے اکٹھے سرطان والے (ایچ۔بی۔سی۔سی HBCC) خاندانوں میں تحقیق کی ہے۔ ہمارے لئے ایچ۔بی۔سی۔سی سرطان والے خاندان اس لئے دلچسپی کا باعث بنے کیونکہ ان خاندانوں کے 1/5 حصہ میں چیک ٹو جین میں *1100delC* والی تبدیلی (*CHEK2 1100delC*) بھی دریافت ہو چکی ہے۔جس کے بارے میں یہ خیال کیا جاتا ہے کہ یہ جین اور دوسرے جینز باہمی شراکت کے ساتھ چھاتی کے موروثی سرطان کا باعث بنتے ہیں۔ہمارے تحقیق کی بنیاد اس حقیقت پر بھی مبنی ہے کہ ایم۔ایس۔ایچ 6 اور چیک۔2 جینز لنچ سنڈروم کا باعث بننے میں اکٹھے منسلک پائے گئے ہیں۔اسی وجہ سے ہم نے 68 ایچ۔بی۔سی۔سی کے سرطان کے خاندانوں کے خون سے موصول ہوئے ڈی این اے (DNA) میں ایم ایس ایچ کا جین کے پروٹین بنانے والے حصہ میں جنیاتی تبدیلیاں معلوم کی ہیں۔ ہم نے دیکھا کہ 11 ایم ایس ایچ 6 جین میں تمام نایاب تبدیلیاں جن کی تعداد ایک فیصد سے کم ہے وہ ایچ۔بی۔سی۔سی خاندانوں میں کنٹرول کی نسبت 11.8 فیصد پائی گئی ہیں۔

لہذا ایم۔ایس۔ایچ 6 جین میں یہ نایاب تبدیلیاں ایچ۔بی۔سی۔سی سرطان کے ساتھ منسلک ہونے کا اشارہ دیتی ہیں (0.00001> پی ویلیو)۔علاوہ ازیں ان تبدیلیوں کی غیر ایچ۔بی۔سی۔سی، (0.001> پی ویلیو) سرطان سے منسلک ہونے کی بھی نشاندہی ہوتی ہے۔جس سے یہ نتیجہ اخذ کیا جا سکتا ہے کہ ان تبدیلیوں کا تعلق صرف ایچ۔بی۔سی۔سی سرطان کے ساتھ نہیں بلکہ غیر ایچ۔بی۔سی۔سی سرطان کے ساتھ بھی ہے۔ یہ سب نتائج بتاتے ہیں کہ ایم۔ایس۔ایچ 6 جین میں رونما ہونے میں والی نایاب تبدیلیاں چھاتی کے موروثی سرطان کا باعث بن سکتی ہیں۔تاہم اب تک ان نایاب تبدیلیوں میں کوئی بھی واضح طور پر خطرناک ثابت نہیں ہو سکی ہے جس کا مطلب ہے کہ ان تبدیلیوں کا حیاتیاتی طریقہ کار چھاتی کے سرطان کے رونما ہونے میں بہت سست ہو سکتے ہیں۔اس کے بعد ہم نے دیکھا کہ چھاتی کے (1) سرطان کے خطرے والے لوسائی کس طرح سے چھاتی کے سرطان کا خطرہ میں بڑھوتری کا باعث بنتے ہیں اور آیا (2) کہ لوسائی چھاتی کے سرطان کے مریضوں میں بیماری کے پھیلاؤ کی رفتار کو متاثر کرتے ہیں یا نہیں۔اس تحقیق کے لئے ہم نے 7 چھاتی کے سرطان سے منسلک لوسائی کا احاطہ کرنے والی 9 امیدوار سنپس (SNPs) کی جنیاتی اقسام دریافت کی۔ان سنپس کی جنیاتی اقسام کی معلومات ہم نے 40 انسانی چھاتی کے سرطان کی سیل لائنز (باب نمبر 4) اور 2480 ابتدائی چھاتی کے سرطان کے مریضوں جن کی مکمل طبعی معلومات حاصل تھی کی رسولیوں (Tumors) میں کی (باب نمبر 5)۔

اس کے ساتھ ساتھ ہم نے چھاتی کے سرطان کے ان لوسائی سے منسلک بیماری کا حیاتیاتی طریقہ کار معلوم کرنے کے لئے ایک ممکنہ مفروضہ پر تحقیق بھی کی۔اس مفروضہ کے مطابق چھاتی کا سرطان

ان لوسائی کے قریب میں موجودہ جینز کے اظہار میں اتار چڑھاؤ کی وجہ سے ہوسکتا ہے۔ ہمارے نتائج کے مطابق ان لوسائی کے اقلیتی ایلیل کی شرح (مائنر ایلیل فریکوئنسی؛MAF) یورپی شجرہ نسب کے تندرست لوگوں کے خون سے موصول شدہ ڈی۔این۔اے کی نسبت سرطان کی سیل لائنز اور مریضوں کی رسولیوں میں زیادہ پائی گئی ہے۔

ہماری تحقیق کے نتائج کے مطابق سنپ آر۔ایس3803662 (SNP rs3803662) جو کہ ٹاکس 3 (*TOX3*) جین کے قریب واقع ہے، اس جین کے کم اظہار کے ساتھ منسلک پائی گئی ہے۔ اس کے علاوہ دیگر چھاتی کے کم خطرہ کے لوسائی سیل لائز اور رسولیوں میں اپنے قریبی جینز کے اظہار پر عمومی طور پر اثر انداز نہیں ہوتی ہیں۔ مزید ہمارے مشاہدے کے مطابق یہ لوسائی بیماری کے پھیلاؤ کی رفتار میں تبدیلی کے ساتھ بھی عمومی طور پر منسلک نہیں ہیں، ماسوائے سنپ ایچ 19 (آر۔ایس2107425.جی/اے (rs2107425G/A) جو کہ ایچ 19 جین (*H19*) کے قریب واقع ہے، اس سنپ کے موجود ہوموزائیگس (Homozygous) خطرے والے ایلیل بیماری کے دوسرے حصوں میں انتقال کی معیاد سے منسلک پائے گئے ہیں۔ اگرچہ یہ چھاتی کے کم خطرے والے لوسائی سرطان کے رونما ہونے کے ساتھ بار بار منسلک کئے گئے ہیں لیکن ہم اپنی تحقیق کی بنیاد پر یہ کہتے ہیں کہ یہ لوسائی عمومی طور پر اپنے قریب میں واقع جینز کے اظہار اور بیماری کے پھیلاؤ میں تبدیلی کا باعث نہیں بنتے (باب نمبر 5)۔

اس مقالہ کے دوسرے حصہ میں ہم نے چھاتی کے سرطان کی سیل لائز کو استعمال کرتے ہوئے سرطان کی مختلف اقسام سے تعلق رکھنے والے حیاتیاتی عناصر مثلاً جینز اور مائکرو آر۔این۔اے (miRNA) پر تحقیق کی ہے۔ مزید ہم نے ان چھاتی کے سرطان کی سیل لائز کے پیوند بالغیر (xenograft) چوہوں میں تیار کئے ہیں اور یہ سمجھنے کی کوشش کی ہے کہ آیا ان مختلف اقسام سے منسلک حیاتیاتی اور طبعی عناصر کا رویہ کیا ہوسکتا ہے اور ان عناصر کے ساتھ نئی اور پہلے سے موجود سرطان کی ادویات کیسے مداخلت کرتی ہیں۔ اس کے علاوہ ہم نے 1427 مریض جن کے طبعی معائنے کا ڈیٹا موجود تھا، ابتدائی رسولیوں کو استعمال کرتے ہوئے ٹوسٹ 1 (*TWIST1*) جو کہ ایک مشہور ٹرانسکرپشن عنصر اور اپی تھیلیم سے میزنیکائمل میں تبدیلی (ای۔ایم۔ٹی؛EMT) کا مارکر ہے، کی بیماری کے پھیلاؤ کی رفتار سے منسلک ہونے کی اہمیت پر تحقیق کی ہے۔ مزید یہ کہ ہم نے پہلے سے موجود جینز کے اظہار کا ڈیٹا استعمال کرتے ہوئے ٹوسٹ۔1 جین کی بنیادی حیاتیات پر تحقیق کی ہے۔ ہمارے نتائج بتاتے ہیں کہ ای۔آر۔پازیٹو (ER+) اور لمف نوڈ نیگیٹو (LN-) چھاتی کے سرطان میں ٹوسٹ 1 جین کا ایم۔آر۔این۔اے (mRNA) کی سطح پر کثیر اظہار بیماری میں پھیلاؤ کی تیز رفتار کی نشاندہی کیلئے ایک مشتند اور آزاد مارکر ہے۔ ہماری تحقیق مزید ثبوت فراہم کرتی ہے کہ ٹوسٹ 1 کے اظہار کے منفی اثرات کا تعلق رسولیوں کے گرد نواح اور ٹشو میں رابطی ریشوں کی کثیر تعداد سے ہوتا ہے (باب نمبر 6)۔

پھر ہم نے یہ جاننے کے لئے کہ انسانی چھاتی کے سرطان کی سیل لائنز ایم۔آر۔این۔اے اور مائیکرو آر۔این۔اے کے اظہار کی سطح پر ابتدائی طبعی رسولیوں کا متبادل ہوسکتی ہیں یا نہیں، 51 سیل لائنز میں ایم آر۔این۔اے اور مائیکرو آر۔این۔اے کے اظہار کا جائزہ لیا (باب 7)۔ ایم۔آر۔این۔اے اور مائیکرو آر این اے کے پروفائلز ان سیل لائنز کو ای۔آر پوزیٹو اور ای۔آر نیگیٹو سرطان میں تقسیم کرتے ہیں۔ یہ مشاہدہ اُس عمومی طور پر قبول شدہ مفروضے سے ملتا ہے، جس کی بناء پر یہ کہا جاتا ہے کہ ای۔آر پوزیٹو اور ای۔آر نیگیٹو سرطان حیاتیاتی طور پر مختلف ہیں۔ ہمارے نتائج مزید بتاتے ہیں کہ کیوں مائیکرو۔آر۔این۔اے کے اظہار میں فرق کا تعلق چھاتی کی سرطان کی مختلف اقسام سے ہے، کچھ کا تعلق سیل لائنز میں ای۔ایم۔ٹی کے مختلف درجات سے ہے اور کچھ کے اظہار میں فرق کا تعلق نارمل سیل لائنز اور ایسی سیل لائنز جن میں رسولیوں کا باعث بننے والے جینز (Oncogenes) اور رسولیوں کی بڑھوتری کو روکنے والے جینز (Tumor supressor genes) تبدیل ہوئے ہوں، سے ہے۔ یہ نتائج ظاہر کرتے ہیں کہ یہ انسانی چھاتی کی سیل لائنز اپنی حیاتیاتی خصوصیات جو کہ رسولیوں میں پائی جاتی ہیں قائم رکھتی ہیں۔ اس لئے ہم کہہ سکتے ہیں کہ سیل لائن کا یہ ماڈل چھاتی کے سرطان کی حیاتیات پر تحقیق کے لئے ایک موضوں ماڈل ہے۔ اس کے علاوہ ہمارے ان سیل لائنز سے موصول ہونے والا مائیکرو آر۔این۔اے کا ڈیٹا منفرد ہے جو کہ چھاتی کے سرطان میں طبعی اہمیت کے حامل مائیکرو آر۔این اے کی حیاتیات کو واضح کرنے کے لئے مفید ثابت ہوسکتا ہے۔

آخر میں ہم نے ان سیل لائن ماڈلز کی اہمیت کو جانچا ہے آیا کہ یہ سیل لائنز چوہوں میں نشوونما پاسکتی ہیں اور اگر ایسا ہے تو کیا ان میں جسم کے دوسرے حصوں میں منتقل ہونے کی اہلیت بھی موجود ہے؟ ہم نے اس بات کا اندازہ بھی لگایا ہے کہ ان ماڈلز کو ابتدائی ذینوگرافٹ ماڈل کے طور پر استعمال کیا جاسکتا ہے اور کیا انکا جسم کے دیگر حصوں میں پھیلاؤ کا رویہ مریضوں میں ابتدائی رسولیوں کی نشوونما کے رویے سے ملتا ہے یا نہیں؟ اس کے لئے ہم نے چھاتی کی ہر ایک قسم میں سے چار سیل لائنز کا انتخاب کر کے ان کو لوسی فیریز کے ایک ریپورٹر جین کے ساتھ دائمی طور پر ضم کیا، اور پھر ان تیار شدہ سیل لائنز کو چوہوں کی چھاتی کے نیچے پیوند کر دیا۔ ہمارے نتائج کے مطابق ان سیل لائنز کی اکثریت (87.5%) چوہوں میں نشوونما پاتے ہوئے کی پیوند یا بغیر (زینوگرافٹ) رسولی بنانے کے قابل تھیں۔ ٹھیک جس طرح انسانی طبعی رسولیوں میں دیکھا گیا ہے اُسی طرح نارمل جیسی/کم کلاڈین (Normal-like/Claudin-low) والی سیل لائن کی قسم جو کہ ای۔ایم۔ٹی اور چھاتی کے سٹم سیل کی کم خصوصیات کی حامل ہے، باقی اقسام کی نسبت بہت تیزی سے چوہوں میں نشوونما پاتی تھیں۔ (مان-وتنی ٹیسٹ، پی= 0.001) سرطان کے جسم کے دیگر حصوں میں پھیلاؤ کو دیکھتے

ہوئے ہم نے مشاہدہ کیا کہ تمام اقسام کی سیل لائنز بغیر کسی امتیاز کے لمف نوڈ، پھیپھڑوں اور جگر کو متاثر کرتی ہیں (باب 8)۔

اس ڈیٹا سے ہم یہ نتیجہ اخذ کرتے ہیں کہ باوجود سال ہا سال لیبارٹی کے مصنوعی ماحول میں نشوونما پانے والی ان انسانی چھاتی کے سرطان کی سیل لائنز نے اپنی آبائی اور قدرتی طاقت کو برقرار رکھتے ہوئے چوہوں میں نہ صرف نشوونما پائی ہے بلکہ جسم کے دیگر حصوں میں بھی منتقل ہوئی ہیں۔ اس لئے ہم کہتے ہیں کہ ان سیل لائنز کو انسانی چھاتی کے سرطان پر وسیع جنیاتی اور حیاتیاتی وجوہات کی تحقیق کے لئے ایک ابتدائی آلے کے طور پر استعمال کیا جاسکتا ہے اس کے علاوہ ان کے زینوگرافٹ کو چھاتی کے سرطان کی ادویات کو ابتدائی طور پر ٹیسٹ کرنے کے لئے استعمال کیا جاسکتا ہے۔

Appendices

Acknowledgements

List of publications

PhD Portfolio

Author's Curriculum Vitae

Acknowledgements

All praises to Almighty Allah, the most gracious and the most beneficent, Who has bestowed me with the strength and courage to finish this PhD thesis successfully. This was indeed a challenging journey. I must admit that the completion of this thesis would not have been possible without the help of many people, who whenever needed, assisted and encouraged me to get the job done. I would therefore like to take a while to acknowledge the contribution of each and every one who facilitated me to achieve this goal.

First of all, I would like to express my sincerest gratitude to Dr. Meike Schutte, who provided me a great opportunity to do my PhD dissertation in her lab. Dear Mieke, you were a great mentor with a fertile brain and above all you were a great person. Your way of schooling to a young PhD student in the lab was incredible, which not only transformed me into a good researcher but also strengthened my theoretical background of cancer genetics. Besides that I enjoyed a lot discussing various cultural and religious aspects with you though those discussions never ended on common grounds. One of the most pleasant parts of being in your lab was going out for the lab day, which I think was unique to your lab. It was sad that you left science during the second year of my PhD. Though it was tough part but it became easy as you left me in the expert hands of John Foekens and John Martens.

My deepest regards go to my promoters, Prof. Jan Klijn and Prof. John Foekens. Dear Jan, please accept my special thanks for being all the time much welcoming and caring person. Dear John thanks a lot for giving me a chance of continuing my PhD in your group. With the departure of Mieke, it would have been difficult for me to start all over. You not only embraced me in your group but also provided me with the financial support, which later on enabled me to bring my family in the Netherlands. Life became much beautiful to have my family with me. Moreover, your reviews of my manuscripts and thesis were very detailed and you always had insightful comments and suggestions. Besides all this, you taught me how important is to keep up morality and ethics in research. I wish you a very blissful life ahead.

I am extremely grateful to my co-promoter, Dr. John Martens for providing me opportunity to work under his supervision. I considered myself to be lucky person to have a supervisor like you, who not only opened the doors of his lab for me but also gave me plenty of independence to set up my experiments. Dear John, you are a great leader, extremely polite and easily approachable supervisor. Without your assistance and advices, I would not have been able to write up successful articles and this thesis. Your warm and welcoming personality has always encouraged me to drop by your office whenever I had something to discuss. Above all, whenever I needed a “smart sentence” during my write-up, I always knew from where to get it. My memories with you span beyond the work. I and other colleagues many times enjoyed your company outside the lab on nice informal drinks and dinners. Indeed, the presence of yours Jenny on those occasions had always been a very pleasant addition. Thank you very much Jenny for being so generous to be with us.

I am grateful to the members of the small committee of my thesis: Prof. André Uitterlinden, Prof. Handrik Marion de Jong, and Dr. Els Berns for their time and energy to review my thesis in such a short time span. Dear Els, I owe you my sincere appreciations. You are a very dynamic and pleasant personality. Your comments and suggestions on my papers and thesis were always constructive and valuable.

The completion of this thesis would not have been possible without the kind assistance of Dr. Anieta Sieuwerts, an impressive researcher and extremely kind person. Dear Anieta, in these years I had the pleasure to discuss many technical and methodological issues with you, which had taught me a lot. You had always been humble and polite in teaching me every now and then, especially our scientific discussions at the lab in the week ends were extremely effective. Most importantly, your advices and comments on my papers were always exceptional and helpful. I would not be fair to forget the support of Joan Bolt. Dear Joan, you were always there to help me whenever I needed especially your help in regard to animal experiments, writing DEC protocol etc. was incredible. Moreover, your help in translating Dutch bureaucratic letters for me and my family was the most favorable part. My kind and special regards also go to Mieke Timmerman and her team for their timely help to accomplish animal xenograft project. Dear Anita thanks a lot for helping me in making all those cytospins and histological pictures, which indeed became the essential part of my thesis.

Good colleagues are always a blessing. In this regard I was lucky to have very lively and supportive companionship of Fons Elstrodt, Antoinette Hollestelle and Marijke Wasielewski. Dear Fons, I am extremely thankful to you. You were my first mentor and an excellent technical person in Mieke's lab. I would not forget your generosity of furnishing my room with some of your personal stuff and helping me out with my initial documentation at the City Hall. Dear Antoinette, you were also a decent, talented and knowledgeable person. Your initial work on human breast cancer cell lines had inspired me to continue working on this model, which indeed resulted into two important chapters of my thesis. I feel honored to pay my sincere gratitude to you for being so kind and helpful colleague throughout my PhD trajectory. Dear Marijke, it was wonderful experience working with you. I learned many aspects of familial breast cancer from you. I wish you all a very successful career in future.

My deepest thanks go to Maxime Look, Marcel Smid, and Anouk Heine (the bioinformatics team of the group). Dear Maxime and Marcel, you never ever said no to any of my 'small' and big questions and without your kind and timely help it would not have been possible to finish this thesis in time. Dear Anouk, you were my first Dutch language teacher and it was because of you I got familiar with all those Dutch words, which I would not had been able to learn on my own.

It is my pleasure to thank those whom I have been collaborating at Erasmus MC during my PhD tenure, Abbas Dehghan (Epidemiology department), Roel Janssen (Genetics department), Marijn van Jaarsveld (Erik Wiemer Lab) and Marleen Malis (Nuclear Medicine department). It had been wonderful working with you all. My thanks also go to Jean Helmijr (little master) for his great artistic touch in cover designing and formatting my thesis. It is always great to have a master like him in

the lab. I would also like to mention Prof. Stefan Sleijfer (Daniel Den Hoed cancer center), Dr. Arzu Umar and Dr. Maurice Jansen for their inspirational scientific discussions and advices.

I pay my sincere gratitudes to late Buddy Hann, who had been mentoring me with my animal experiments even at the last moment of his life. May his soul rest in peace! Many thanks also go to the people of animal facility of Erasmus MC for their technical help and support.

My special thanks go to my office mates and PhD fellows, Bahar, Mila, NingQing (my paranymph), Katherina, Diana and Jingjing for providing me a great multicultural environment. Besides, I would also like to appreciate all others PhD and Posdoc fellows, René Braakman, Tomasso, German, Raquel and Gaia for their very warm, cheerful, and international company, which had always been incredibly pleasant. I am also grateful to Wendy Pragersmissen, Vania, Renee Foekens and Kirston for their technical support.

Dear Dana Mustafa, You were a special person for me and my family. I would never ever forget your unconditional help and care especially at the time of Ayaan's birth (my elder son). I wish you a very successful career in future. Dr. Muhammad Akram and Mesha Saeed, thanks a lot for helping me for Urdu translation of the summary of this thesis.

Many sincere thanks to Shuakat Bhati who had been the most prominent and charming fellow for me in both Pakistan and The Netherlands. I wish him a great career ahead. My deep gratitude goes to Naseer uncle and his family, Riaz Bhai and his family and Haji sab for being so much supportive, caring, and loving during our stay in the Netherlands. I wish that Allah (S.W.T) always shower His special bounty and blessings upon them. Good wishes and special thanks also go to Aunty Sameena and Aunty Banu for providing me and my family a kind hospitality in the Netherlands.

My friends are my asset and I would never ever forget their absolute love, care and sensational company. Dear Riaz, Zafar (Balouch), Dr. Akram (PU), Dr. Akhtar (Wasim), Dr. Khan (Altaf), Shoaib (my paranyphms), Aamir, Malik Saboor, Talal Gondal, Naveed (Sweden), Raju (India) and many more, you people are just great. One cannot lose anything in life to have you incredible guys on the back. I really wish you all the best in your life.

This journey began with the persuasion from a sweet peer, Sharif Rana (Chaudhary sahb). His encouragement lit a spark in me to go for higher studies. I wish the best for him and his family.

Finally, I pay my humble tributes to my parents for their countless prayers, everlasting love, sincere care and support. My parents are the real reason behind my success. I am grateful to my brother, Muhammad Razzak and my sisters for sharing my happiness and sorrows and providing me support through thick and thin. Finally, special thanks to my wife and my sons Muhammad Ayaan Riaz, Muhammad Farhan Riaz who gave me courage and supported me to complete this thesis and also for being loving, and affectionate and providing me with an ideal environment to work even at the cost of their own comfort.

LIST OF PUBLICATIONS

Publications related to this thesis

Growth and metastatic behavior of human breast cancer cell lines in mice.
Muhammad Riaz, Buddy Setyono-Han†, Mieke A. Timmermans, Anita, M. Trapman, Joan Bolt, Antoinette Hollestelle, Maxime P. Look, John A. Foekens, and John W.M. Martens.
In preparation

MicroRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs.
Muhammad Riaz, Marijn T.M. van Jaarsveld, Antoinette Hollestelle, Wendy J.C. Prager-van der Smissen, Anouk A.A.J. Heine, Antonius W.M. Boersma, Jingjing Liu, Jean Helmijr, Bahar Ozturk, Marcel Smid, Erik A. Wiemer, John A. Foekens, and John W.M. Martens.
Submitted for publication

High *TWIST1* mRNA expression is associated with poor prognosis in lymph node-negative and estrogen receptor-positive human breast cancer and is co-expressed with stromal as well as ECM related genes.
Muhammad Riaz, Anieta M. Sieuwerts, Maxime P. Look, Mieke A. Timmermans, Marcel Smid, John A. Foekens, John W.M. Martens.
Breast Cancer Res. 2012; 14(5):R123

Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival and mRNA expression of the nearest genes.
Muhammad Riaz, Els M.J.J. Berns, Anieta M. Sieuwerts, Kirsten Ruigrok-Ritstier, Vanja de Weerd, Arwin Groenewoud, Andrè G. Uitterlinden, Maxime P. Look, Jan G.M. Klijn, Stefan Sleijfer, John A. Foekens, and John W.M. Martens.
Breast Cancer Res Treat. 2012; 133(3):843-51

Association of rare MSH6 variants with familial breast cancer.
Marijke Wasielewski, **Muhammad Riaz,** Joyce Vermeulen, Ans van den Ouweland, Ineke Labrijn-Marks, Renske Olmer, Linda van der Spaa, Jan G.M. Klijn, Hanne Meijers-Heijboer, Dennis Dooijes, and Mieke Schutte.
Breast Cancer Res Treat. 2010; 123(2):315-20

Low-risk susceptibility alleles in 40 human breast cancer cell lines.
Muhammad Riaz, Fons Elstrodt, Antoinette Hollestelle, Abbas Dehghan, Jan G.M. Klijn, and Mieke Schutte.
BMC Cancer 2009; 9:236

Other publications

Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines.
Antoinette Hollestelle, Jord H.A. Nagel Marcel Smid, Suzzane Lam, Fons Elstrodt, Marijke Wasielewski, Ser Sue Ng, Pim J. French, Justine K. Peeters, Marieke J. Rozendaal, **Riaz M,** Ellen, C. Zwarthoff, Amina Teunisse, Peter J. van der Spek, Jan G.M. Klijn, Stephan P. Ethier, Hans Clevers, Aart G. Jochemsen, Michael A, Bakker, John A. Foekens, John W.M. Martens, and Mieke Schutte.
Breast Cancer Res Treat 2010; 121(1):53-64.

Influence of medium components and physical factors on vitamin B_{12} production by *Pseudomonas sp.* PCSIR 99.
Muhammad Riaz, Zumir A. Ansari, Fouzia Iqbal, Ijaz Ali, Tauqir Ahmed, and Muhammad Akram.
Pak. J. Sci. Ind. Res 2007; 50: 401-407.

Microbial production of vitamin B_{12} by methanol utilizing strain of *Pseudomonas species*
Muhammad Riaz, Zumir A. Ansari, Fouzia Iqbal, and Muhammad Akram.
Pak. J. Biochem. Mol. Biol 2007; 40: 5-10.

Phytosciological survey of ground flora in Chichawatni Forest.
Muhammad Akram, Kalid F. Akbar and **Muhammad Riaz,**
Pak. J. Sci 2007; 59: 110-115.

PHD PORTFOLIO SUMMARY

Name PhD student	Muhammad Riaz
Erasmus MC Department	Medical Oncology
Research School	Molecular Medicine
PhD period	September 2007 to September 2012
Promotors	Prof. dr. J. A. Foekens and Prof. dr. J. G. M. Klijn
Supervisor	dr. J. W. M. Martens.

PhD training	**Year**	**ECTS**
General and specific courses		
Course Basic and Translational Oncology of the Erasmus MC Post Graduate School	2007	1.12
Course Analysis of Microarray Gene Expression Data	2008	1.00
Course Classical Methods for Data analysis	2008	5.70
DNA damage and cancer risk	2009	1.60
SNPs and human disease	2009	1.50
Laboratory animal science	2009	3.00
Biomedical English writing and communication	2010	4.00
In vivo imaging: from cell to organism	2010	1.50
Seminars and workshops		
Monthly JNI Oncology Lectures	2008–2012	1.00
JNI Scientific New Year Symposium	2008–2012	2.50
Annual Medical Oncology Department Meeting	2008–2012	2.50
Weekly JNI Scientific Meetings	2008–2012	2.50
Journal Club		1.00
Annual Molecular Medicine Day in Rotterdam	2008–2012	2.00
Annual CGC Scientific Meeting	2010–2011	0.60
Workshop on Mouse Model at Leiden University	2010	1.50
Presentations		
Journal Club (Presentations)	2008-2012	1.50
Annual Oral Presentation at the JNI Scientific Lab meetings	2008-2011	1.50
MGC Annual Meeting Klohn, Germany (Poster Presentation)	2010	0.90
Annual CGC Meeting on Molecular Mechanisms in Cancer (Poster Presentation)	2010	0.90
33rd Annual CTRC-AACR San Antonio Breast Cancer Symposium – in San Antonio, Texas, USA (Poster Presentation)	2010	1.00
16th ECCO-ESMO Stockholm Cancer Conference – in Stockholm, Sweden (oral presentation)	2011	2.00
16th Molecular Medicine Day, Erasmus MC (poster presentation)	2011	0.90
16th Molecular Medicine Day, Erasmus MC (poster presentation)	2012	0.90
(Inter)national conferences		
San Antonio Breast Cancer Symposium 2010 in Texas in USA	2010	1.00
ECCO conference in Sweden	2011	1.00
Total ECTS		**44.62**

AUTHOR'S CURRICULUM VITAE



Muhammad Riaz was born in Okara, a splendid city in Pakistan on February 15, 1979. He passed his master in Applied Chemistry from University of Engineering and Technology, Lahore, Pakistan in 2001. Being much fascinated by the role of nucleic acid in biology, he went to study Molecular Biology and did two year M. Phil degree in Molecular Biology from National Center of Molecular Biology, University of the Punjab, Lahore. In 2007, he was awarded a prestigious overseas scholarship for PhD studies by the Higher Education Commission of Government of Pakistan. He started his PhD research in the Laboratory of Oncogenetics under supervision of Dr. Mieke Schutte in the Department of Medical Oncology, Erasmus University Medical Center – Daniel den Hoed Cancer Center. After the departure of Dr. Schutte, Muhammad continued working in the department on genomics in breast cancer under direct supervision of Dr. John Martens in the research section "Translational Cancer Genomics and Proteomics" headed by Prof. Dr. John Foekens. During five years of his PhD tenure, Muhammad worked on various aspects of breast cancer, which included analysis of genetic and prognostic factors in breast cancer and *in vitro* and *in vivo* studies on breast cancer subtypes. The results of his research have been presented on various international scientific meetings including the San Antonio Breast Cancer Symposium and the European CanCer Organization (ECCO), the latter involved an oral presentation at the main stage. His PhD research has led to three published scientific articles and two research articles submitted for publication as a first author and two articles published as co-author.