Significance of *WRAP53* in Breast Cancer

Mutation Analyses and Gene Expression Studies

by

Veronica Okkenhaug Vang

2012

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Significance of *WRAP53* in Breast Cancer

Mutation Analyses and Gene Expression Studies

by

Veronica Okkenhaug Vang

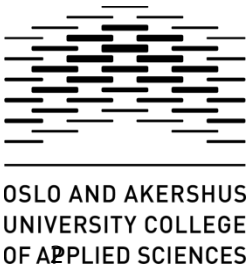
Master program in Biomedicine

Department of Health Sciences

Thesis submitted for the Master`s degree in Biomedicine, 60 ECTS, Oslo University Hospital, Institute for Cancer Research, Department of Genetics and Oslo and Akershus University College of Applied Sciences

Supervisor: Dr. Anita Langerød

May 21st, 2012



Acknowledgements

The work presented in thesis is carried out at the Department of Genetics, Institute for Cancer Research, the Norwegian Radium Hospital from August 2011 to May 2012 for the Master`s degree in Biomedicine at Oslo and Akershus University College of Applied Sciences.

I would like to thank Professor Anne-Lise Børresen Dale for welcoming me into her group, for including me into the everyday activities at the department and for providing me with top quality laboratory facilities during my stay. Special thanks go to my supervisor Dr. Anita Langerød for all her friendliness, time and invaluable suggestions and advices. Your professional dedication and enthusiasm is inspiring.

I will express my gratitude to Phuong Vu, Eldri Undlien Due and Anita Halvei for sharing their laboratory expertise, and for always answering my questions with a smile. Thanks to Laxmi Silwal-Pandit for sharing her knowledge about *WRAP53* and mutation analyses, and to Anja Valen for patiently teaching me GeneSpring. I would also like to thank all the fabulous Genetics people for creating such an including and inspiring working environment, it has been great!

A special thank goes to the people at the Department of Oncology and Pathology, Cancer Centre Karolinska, Karolinska Insitutet, Dr. Marianne Farnebo, Dr. Adam Clauss and Dr. Elisabeth Hedström, for cooperation in the gene expression study. Thanks for all your efforts concerning planning and performance, for taking good care of our cells, and for your quick response to all my questions. It is very much appreciated.

Thanks to the Comprehensive Cancer Center Council, the Norwegian Radium Hospital Research Foundation and Helse Sør-Øst for financial support.

Finally, I would like to thank friends and family for all their nice encouragements. My heartfelt gratitude goes to my dear Jon for his endless patience and support, excellent computer aid, and for reminding me to enjoy science even though it has been some demanding months.

The Norwegian Radium Hospital, Oslo, 18th of May 2012
Veronica Okkenhaug Vang

Abstract

WRAP53 partly overlaps the neighboring gene *TP53* in a head-to-head fashion. The *WRAP53α* isoform was identified as a natural antisense transcript to *TP53*, possessing regulatory functions involved in the induction of cellular p53 responses [1]. *WRAP53* encodes as well a protein discovered as a telomerase holoenzyme subunit important for proper telomere synthesis [2]. *WRAP53* overexpression is frequently observed in cancers, presuming *WRAP53* as a gene implicated in cancer cell survival [3]. The overall aim of this thesis has been to investigate the significance of and increase the knowledge about *WRAP53* in breast cancer. This was accomplished performing a *WRAP53* mutation analysis and a gene expression study analyzing cellular responses to *WRAP53* knockdown.

The *WRAP53* mutation analysis was performed sequencing all ten *WRAP53* coding exons and start exon 1β in 175 primary breast carcinomas. Sequence alterations detected in tumor were verified in corresponding blood samples and investigated in relation to clinical, pathological and molecular parameters. The gene expression study was performed inducing siRNA-mediated *WRAP53* depletion in the MCF-7 and MDA-MB-231 breast cancer cell lines. The gene expression alterations caused by *WRAP53* knockdown were analyzed by microarray technology and related to biological functions including cellular pathways and disease states.

No distinctive somatic alterations were detected in the *WRAP53* mutation analysis, indicating that *WRAP53* mutations do not seem to be a common event in breast tumorgenesis. Contrarily, the genetic linked polymorphic alterations, R68G, F150F and A522G, were significantly associated to breast cancer-specific survival and exon 1β c.–245 G>C was associated with nuclear WRAP53 protein localization, a feature of favorable prognostic impact. Together these results suggest that *WRAP53* might be a marker of prognostic value in breast cancer. *WRAP53* depletion altered the gene expression patterns in MCF-7 and MDA-MB-231. Despite cell line-specific outcomes in the pathway analyses, relating differentially expressed genes to biological functions, cancer and cancer-related features like cellular proliferation, growth and movement emerged as significant common denominators. The same was observed analyzing the mutual differentially expressed MCF-7 and MDA-MB-231 genes, supporting the hypothesis that *WRAP53* might be involved in tumorgenesis.

Sammendrag

WRAP53 og *TP53* er delvis overlappende gener, og isoformen *WRAP53α* er identifisert som et naturlig antisens transkript til *TP53* med regulatoriske funksjoner involvert i aktivering av cellulær p53-respons [1]. Proteinet *WRAP53* koder for er identifisert som en subenhet i enzymkomplekset telomerase, og spiller en avgjørende rolle i telomersyntesen [2]. Overuttrykk av *WRAP53* forekommer ofte i kreft og studier indikerer at *WRAP53* er nødvendig for kreftcellenes overlevelse [3]. Hensikten med denne studien har vært å undersøke betydningen av og øke kunnskapsnivået om *WRAP53* i brystkreft. Dette ble utført ved en *WRAP53* mutasjonsanalyse og en genekspresjonsstudie for å studere cellulære responser av å slå ut *WRAP53* (såkalt "knockdown").

I mutasjonsanalysen ble de ti kodende *WRAP53* exonene og start-exon 1β sekvensert i 175 primære brystkarsinomer. Sekvensvariasjon detektert i tumor ble validert i korresponderende blodprøver, og utforsket med hensyn til kliniske, patologiske og molekylære parametere. Ekspresjonsstudien ble utført ved siRNA-indusert *WRAP53* knockdown i brystkreft cellelinjene MCF-7 og MDA-MB-231. Endringer i genuttrykk som følge av *WRAP53* knockdown ble analysert ved hjelp av microarray teknologi og relatert til biologiske funksjoner som cellulære signalveier og patologiske tilstander.

Ingen karakteristiske somatiske endringer ble påvist i mutasjonsanalysen, noe som indikerer at mutasjoner i *WRAP53* trolig ikke er hyppig forekommende hendelser i brystkreft. De genetisk koblede polymorfiene, R68G, F150F og A522G, ble funnet signifikant assosiert med brystkreft-spesifikk overlevelse, og exon 1β c.–245 G>C ble assosiert med kjernelokalisering av WRAP53 proteinet, noe som bedrer prognosene ved brystkreft. Disse resultatene indikerer at *WRAP53* kan være en markør med prognostisk verdi ved brystkreft. Knockdown av *WRAP53* endret genuttrykket i MCF-7 og MDA-MB-231. Til tross for celletype-spesifikke resultater ved analysering av de signifikant endrede ekspresjonsnivåene i assosiasjon til biologiske funksjoner, ble kreft og kreft-relaterte karakteristikker som cellulær vekst, proliferasjon og bevegelse angitt som signifikante fellestrekk. Det samme ble observert ved analyse av gener med endret ekspresjonsnivå felles uttrykt i MCF-7 og MDA-MB-231, noe som støtter hypotesen om at *WRAP53* kan være involvert i kreftutvikling.

Abbreviations

Ala (A)	Alanine
Arg (R)	Arginine
ATF2/3	Activating transcription factor 2/3
BC	Base change
B-H	Benjamini-Hochberg
bp	Base pair
BRCA1/2	Breast Cancer 1, early onset/Breast Cancer 2, early onset
CB	Cajal body
CCK	Cancer Center Karolinska
cDNA	Complementary DNA
cRNA	Complementary RNA
Cy3	Cyanine 3
CYP2D6	Cytochrome P450 2D6
DC	Dyskeratosis congenita
DCIS	Ductal carcinoma *in situ*
dNTP	Deoxynucleotide triphosphate
ddNTP	Dideoxynucleotide triphosphate
DKCB3	Dyskeratosis congenita autosomal recessive type 3
DMT	Dimethoxytrityl
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
EDTA	Ethylenediaminetetra-acetic acid
ER	Estrogen receptor
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
ESR1	Estrogen receptor 1
FDR	False discovery rate
FLJ10385	Full-Length Long Japan 10385
FOSL2	FOS-like antigen 2
FOXL2	Forkhead box L2
GATA3	GATA binding protein 3
Gly (G)	Glycine
HER2	Human Epidermal Growth Factor Receptor 2
His (H)	Histidine
HW	Hardy-Weinberg
IARC	International Agency for Research on Cancer
IHC	Immunohistochemistry
IL-17	Interleukin 17
IL-17A	Interleukin-17A
INF	Interferon
IPA	Ingenuity Pathways Analysis
iPr	Isopropyl
IRF9	Interferon regulatory factor 9
JAK1	Janus kinase 1
LCIS	Lobular carcinoma *in situ*
LD	Linkage disequilibrium
Leu (L)	Leucine
LFS	Li-Fraumeni Syndrome
Lys (K)	Lysine
MAPK	Mitogen-activated protein kinases
Me	Methyl

NAT	Natural antisense transcript
NCBI	National Center for Biotechnology Information
NFκB	Nuclear factor-κB
NR1D1	Nuclear receptor subfamily 1, group D, member 1
NR3C1	Nuclear receptor subfamily 3, group C, member 1
NTC	Non-treated control
p53	Cellular tumor antigen p53
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
Phe (F)	Phenylalanine
PIK3CA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide
PR	Progesterone receptor
Pro (P)	Proline
PTEN	Phosphatase and tensin homolog
qPCR	Quantitative PCR
RB1	Retinoblastoma 1
RELA	V-rel reticuloendotheliosis viral oncogene homolog A
RNP	Ribonucleoprotein
rs number	Reference SNP number
RXR	Retinoic Acid X Receptor
SAM	Significance Analysis of Microarrays
SAP	Sense/antisense pair
scaRNA	Small Cajal body specific RNA
siC	Short interfering RNA control
siRNA	Short interfering RNA
siWRAP53#2	siRNA targeting *WRAP53* exon 2
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
STAT1/2	Signal transducer and activator of transcription 1/2
TAE buffer	Trizma® base/Acetic acid/EDTA disodium salt buffer
TCAB1	Telomerase Cajal body protein 1
TERC	Telomerase RNA component
TERT	Telomerase reverse transcriptase
TLR	Toll-like receptors
T_m	Melting temperature
TP53	Tumor protein p53
TP63	Tumor protein p63
TYK2	Tyrosine kinase 2
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
ULL	Ullevål University Hospital
ULL-B	Blood samples from the Ullevål cohort
ULL-T	Tumor samples from the Ullevål cohort
UV	Ultraviolet
VDR	Vitamin D Receptor
WDR79	WD repeat domain 79
WRAP53	WD repeat containing, antisense to *TP53*
WD	Tryptophan (W) and aspartic acid (D)
wt	wild type

List of contents

1. Introduction

1.1. Cancer

1.1.1. Cancer etiology

Cancer is a generic term used to describe a heterogeneous group of nearly 200 different cancer types [4]. The development and behavior of cancers differs greatly, but the common cancer features are defined as unlimited cellular replication potential and invasion of foreign body tissues in the process towards metastatic disease [5].

Cancer is today one of the most common human diseases worldwide, with persistently increasing incidence rates [6, 7]. This is observed in both economically developed and developing countries, and is primarily caused by population aging and growth. An important secondary cause is the expanding global adaption to cancer-causing behaviors like smoking, alcohol consumption, physical inactivity and obesity [6]. A study by the Norwegian Cancer Society shows that as many as 30–40% of all cancer cases may be prevented by positive lifestyle changes [8]. In developing countries, late diagnosis and limited access to timely, standardized treatments are also contributors to the increase in cancer incidence. Preventive tools such as programs improving the cancer knowledge in the public, early detection and treatment will be of importance to reduce the growing incidence [6].

The latest cancer statistics from the International Agency for Research on Cancer (IARC), GLOBOCAN 2008[1], reported 12,7 million new cancer cases and 7,6 million cancer deaths worldwide in 2008. Lung cancer is the most common cancer overall and is also the most frequent neoplasm in men, while breast cancer occur most frequently in women [6, 7]. In Norway, the Cancer Registry of Norway reported 27520 new cancer cases in 2009, ranging prostate cancer as the most frequent cancer in men and breast cancer as the most frequent in women [9].

1.1.2. Cancer development

Cancer was for a long time assumed to be a disease of environmental origin. As early as in 1914, Theodor Boveri suggested that malignant tumors originated from cellular genomic abnormalities [10], but the importance of cancer genetics did not really

[1] GLOBOCAN 2008: Estimates of cancer incidence, mortality and prevalence worldwide in 2008, International Agency for research on Cancer, World Health Organization; globocan.iarc.fr

emerge until the 1980s. Today cancer is known as a genetic disease characterized by a stepwise cellular accumulation of genetic and epigenetic alterations, but the consideration of cancer as an environmental disease is still standing [11]. Despite the underlying genetic causes, only 10% of all cancer cases occur in people with a family history of cancer, leaving the majority to be of sporadic occurrence [12].

Cancer development is a highly debated subject and several possible theories exist. A well-accepted model is the clonal evolution theory where neoplasms are presumed to develop from a single somatic cell. Accumulation of genomic aberrations over time provides selective cellular advantages promoting the cell to become a growing mutant clone [5, 13]. Histopathological observations of cancer progression supports this theory [13], even though the polyclonality observed in many tumors remains a contradiction to the model [14].

Over the last years, a supplement to the clonal evolution theory has been described based on the possibility of interclonal cooperation as an explanation of tumor polyclonality. This model do not require a single cell to obtain every genomic aberration necessary to reach malignancy, but indicates that partially transformed cells may cooperate in the process towards a malignant phenotype [14, 15]. Cooperation between co-dominant clones are presumed to occur in three possible ways (figure 1); (B) between a pre-malignant cell and a mutant daughter cell, (C) between two clones generated from the same pre-malignant precursor cell, or (D) between two pre-malignant clones generated from independent normal cells. These cooperative mechanisms generate a malignant phenotype faster than what is possible in the lineal evolution [14].

A competing theory in carcinogenesis is the cancer stem cell hypothesis, suggesting that human tumors develop from mutated normal stem or progenitor cells. The stem cell characteristics of self-renewal, indefinite replication potential and differentiation into diverse cell types are abilities highly relevant to malignancy.

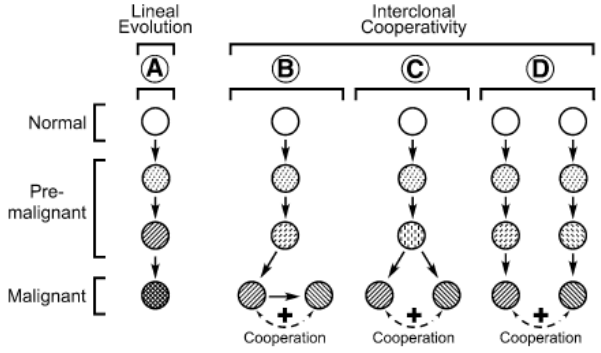


Figure 1: Schematic representation of malignant tumor development by lineal evolution (A) and interclonal cooperativity (B, C and D) [14].

Studies have shown that cells with stem-cell properties are able to induce human tumors, and that such cells have been detected in the hematopoietic system, the central nervous system and the mammary gland [16].

Recently, the hypothesis of cancer self-seeding was introduced, presuming that tumor clones may leave the primary tumor by extravasation into the circulatory system, temporary develop at a distant site, and then return to the primary site generating new subpopulations [17]. Evidence shows that carcinogenesis is a complex process where several theories may be involved in a mutual non-exclusive way, indicating that human cancers may develop differently from each other [13, 14, 16, 17].

Cancer may originate from any body tissue, resulting in extreme disease heterogeneity. Despite this complexity, most human tumors share some common features necessary to reach the state of malignancy. These features are by Hanahan and Weinberg referred to as the "hallmarks of cancer", and constitutes six biological capabilities acquired in cancer cell survival, proliferation and metastatic dissemination; (i) sustain proliferative signaling, (ii) evade growth suppressing signals, (iii) enable replicative immortality, (iv) resist apoptosis, (v) induce angiogenesis and (vi) activate invasion and metastasize. Two emerging hallmarks have during the last decade been added to the list, proving that both evasion of immune destruction and energy metabolism reprogramming are tumor facilitating actions. The cellular hallmark achievements are made possible by two enabling characteristics, where (i) genomic instability increases the cellular mutation rates and contribute to the gain of hallmark capabilities, and (ii) tumor-promoting inflammations recruits and utilize immune cells in the process of tumorgenesis. The tumor microenvironment is another element important in cancer development, showing that cancer progression also depends on a diversity of normal cells. From the early statements describing tumors as homogenous groups of cells, considerable complexity has evolved in the field of cancer biology [18].

1.1.3. Genetic alterations in cancer

Tumor development is primarily caused by genetic and epigenetic alterations in members of two broad gene categories; the proto-oncogenes and the tumor suppressor genes. Proto-oncogenes normally promote cell proliferation and survival while the tumor suppressors restrain cellular proliferative activity, together balancing the

homeostasis of adult body tissue. The genetic alterations are usually of somatic occurrence, but germline mutations may increase cancer susceptibility by familial heritage [19, 20].

Proto-oncogenes are altered by monoallelic gain-of-function mutations generating excessively active oncogenes. Protein products of oncogenes are classified in six categories; transcription factors, growth factors, growth factor receptors, signal transducers, chromatin remodelers and apoptosis regulators, all potential contributors to uncontrolled proliferation if overexpressed. The transition from proto-oncogenes to oncogenes is facilitated by gene amplifications, juxtapositions to regulatory elements of enhanced activity, and structural alterations caused by mutations and gene fusions. In human epidermal growth factor receptor 2 (HER2) positive breast cancer amplification of the *ERBB2* proto-oncogene sustains excessive proliferative signaling [19, 20].

In contrast to the activated oncogenes, tumor suppressor genes are often inactivated in cancer. Biallelic alterations like point mutations, deletions, epigenetic promoter silencing by methylation and loss of heterozygosity result in gene loss-of-function (figure 2). Tumor suppressor gene products inhibit improper cell cycle progression and maintain genomic stability by ensuring proper DNA replication, repair and segregation, and by inducing apoptosis in aberrant cells. Inactivation of such genes becomes a source to increased cellular proliferation and reduced genomic integrity, a phenomenon observed by the frequent inactivation of the *RB1* and *TP53* tumor suppressors in cancer [11, 19].

Tumor genomic instability is caused by inherited mutations in genes important to genomic integrity, or by accumulation of mutations of somatic origin during tumorgenesis [11]. The genetic alterations belongs to different classes of distinctive

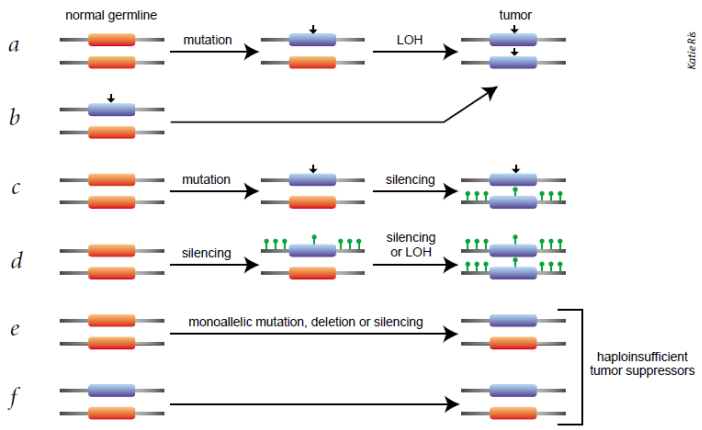


Figure 2: Loss of tumor suppressor gene function in cancer. Function may be lost (blue bars) in germ-line or somatically induced by mutations (vertical arrowheads), loss of heterozygosity or epigenetic promoter methylations (green spikes). For some tumor suppressor genes inactivation of one allele is sufficient for loss of function; the haploinsufficient tumor suppressors [11].

DNA sequence changes including single nucleotide substitutions, insertions, deletions, chromosomal rearrangements and copy number gains and losses [21]. Numerically, single base substitutions are the most abundant genetic variation, causing nonsense, missense or silent mutations. In nonsense mutations the substitutions introduce a premature stop codon and generate truncated, non-functional gene products, while the missense mutations alter the coding sequence by amino acid changes. Some base changes occur as synonymous, silent mutations where the amino acid is retained throughout the substitution [19].

Another emerging field to investigate is the importance of single nucleotide polymorphisms (SNPs) in tumorgenesis. SNPs are single nucleotide alterations defined as normal sequence variations that occur in more than 1% of the population. SNPs are rarely disease causing, and high-frequency SNPs are not expected to have any major phenotypic effect since harmful lesions mainly are eliminated by the process of natural selection. In cancer, SNPs have potential as disease indicators and are used in cancer predisposition analysis, diagnosis and prognosis [19, 22].

Every genetic abnormality is not involved in cancer development. Mutations are classified as “drivers” and “passengers” based on their contribution in cancer development. Driver mutations confer cellular growth advantages, a known oncogenic feature, while the passenger mutations do not provide this advantage and therefore do not contribute to tumorgenesis. An important subgroup of driver mutations is the mutations causing cancer therapy resistance, and as a consequence increased risk of relapse. In cancer genetics, an important goal is to identify the cancer genes carrying driver mutations. Driver mutations are assumed to cluster in genes important in tumorgenesis, while the passengers are distributed more randomly. The identification of cancer genes is difficult since drivers has to be distinguished from passenger mutations, and because multiple low-penetrance cancer genes are believed to contribute in cancer development [21].

Epigenetic alterations are additional contributors to cancer development. Without changing the DNA sequence, only modifying it, epigenetic alterations promotes gene expression changes. Chromatin structure remodeling and DNA methylation are the main mechanisms to these alterations [19].

1.2. Breast cancer

1.2.1. Breast cancer etiology

Breast cancer is the most frequent neoplasm among women worldwide, and ranks second after lung cancer as the most common cancer overall. GLOBOCAN 2008 reported 1,38 million new female breast cancer cases in 2008 [7], a number constituting 23% of all new cancer cases and 11% of all cancer cases in the world [6, 7]. In Norway a total of 2760 persons, distributed in 2745 women and 15 men, were diagnosed for breast cancer in 2009 [9].

Similar to cancer incidence rates in general, the world breast cancer incidence has increased over the last decades. Western European women are burdened with the highest rates (89,9 per 100 000 women), while the lowest are found in Eastern African women (19,3 per 100 000 women). Although the incidence rates overall are higher in economically developed than developing regions, about half of the diagnosed cases and 60% of the deaths occur in developing regions [6, 7]. Breast cancer caused 458.000 deaths in 2008 and is the leading cause of cancer death among women worldwide [6].

Breast cancer is a multifactorial disease caused by interactions of genetic and environmental risk factors [23]. Familial breast cancer history is the strongest predisposing factor despite the fact that inheritance only constitutes about 10% of all breast cancers. Inherited mutations in high-penetrance cancer susceptibility genes like *BRCA1*, *BRCA2* and *TP53* account for about 25% of the familial risk, leaving the remaining 75% of other genetic or environmental origin [11, 24]. Few high-penetrance susceptibility genes have been identified, indicating that most familial breast cancers may be caused by alterations in multiple low-penetrance genes [25].

Endogenous estrogen exposure is a major contribution to breast cancer risk. Increased exposure by early age at menarche, nulliparity, late age at first full-term pregnancy and late age at menopause are associated to increased breast cancer risk. Breast cancer incidence is age-specific and correlates strongly with age until menopause, indicating that ovarian activity is a breast cancer promoting factor. Exogenous hormone therapy like oral contraceptives and postmenopausal hormone replacements are other risk increasing elements.

International variations in breast cancer incidence suggest that environmental and lifestyle factors may influence the risk of breast cancer. High fat diets and alcohol

consumption are breast cancer risk factors as well as postmenopausal obesity, which also increase breast cancer risk. Mammographic density and exposure to ionizing radiation are also well-known risk factors to breast cancer [23].

1.2.2. Breast cancer development and progression

Breast cancer is malignant neoplasms originating from breast tissue. The adult female breast is an intricate organ consisting of glandular tissue surrounded by stromal components like adipose and connective tissue, blood and lymphatic vessels [23]. The mammary gland consists of 15–20 lobes, where each lobe is composed of many smaller lobules and is converged onto a lactiferous duct (figure 3). Secretory luminal epithelial cells face the hollow lumen in the branched ductal and lobular system, while underlying contractile myoepithelial/basal cells facilitates the glandular secretion [26]. The intricate normal breast anatomy contributes to the heterogeneity and complexity observed in breast cancer.

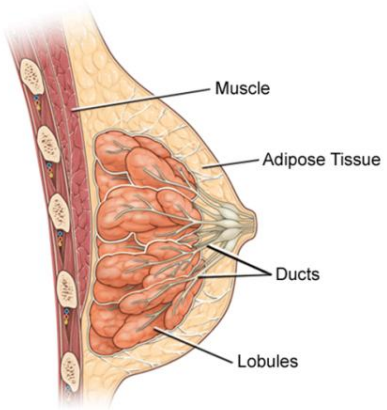


Figure 3: Normal breast anatomy.
The mammary gland consists of lobules converging onto lactiferous ducts surrounded by stromal components like adipose tissue. The pectoralis major muscle separates the breast from the ribs. Modified from http://medicalcenter.osu.edu/greystone/images/ei_0385.gif

Invasive breast cancer is proposed to develop through an evolutionary multistep process from pre-existing benign lesions [27, 28], driven by genetic, epigenetic and microenvironmental changes [24]. Atypical hyperplasia is presumed to be an early step in breast cancer development, later evolving into ductal and lobular carcinomas *in situ* (DCIS and LCIS) (figure 4). Hyperplasia and *in situ* carcinomas acquire certain malignant properties like uncontrolled cell proliferation, but are considered premalignant because of their inability to invade and metastasize [28].

Not every premalignant lesion develops into invasive breast cancer, but patients diagnosed with atypical hyperplasia and *in situ* carcinomas are more susceptible to invasive breast cancer progression [28]. Invasive cells brakes through

the basal membrane of carcinoma *in situ* and invade the surrounding tissue. Myoepithelial/basal cells are lost in the transition from *in situ* carcinoma to invasive cancer, and it is hypothesized that this transition is facilitated by the myoepithelial/basal cells themselves [24]. Invasive ductal and lobular carcinomas are the most frequent types of breast cancer, representing about 80% and 10% of all diagnoses respectively [27], originating from epithelial cells in the terminal ductal lobular units [28]. Metastasis is the lethal aspect in breast cancer, where cells escape from the primary tumor site and spread to distant organs generating secondary tumors [24]. Even though the clonal evolution theory is well established in breast tumorgenesis, hypothesis like cancer self-seeding, cancer cell cooperation and mammary cancer stem-like cells propose different possible explanations to breast cancer development [14, 16, 17]. All tumors may therefore not go through every step of the model shown (figure 4).

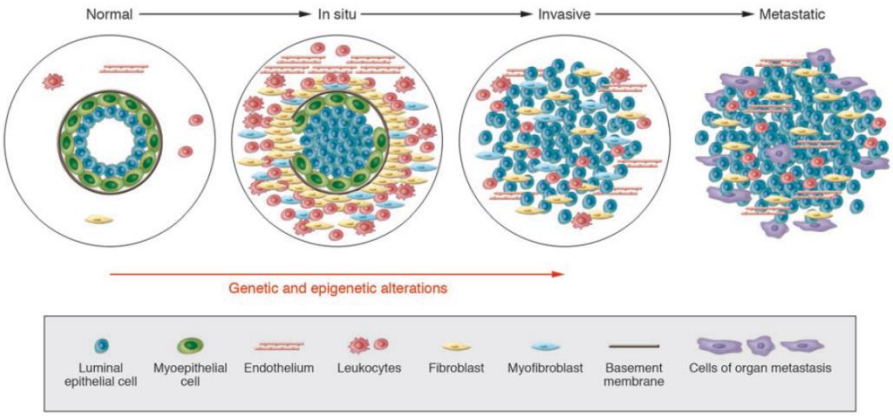


Figure 4: Hypothetical breast cancer progression model progressing through the stages from normal to *in situ*, invasive and metastatic carcinoma. The stromal compartment includes leukocytes, fibroblasts, myofibroblasts, and endothelial cells [24].

1.2.3. Breast cancer subtypes

Over the last decade, an increased molecular approach to breast cancer diagnostics has evolved. The molecular heterogeneity observed in human cancers was primarily found to be explained by transcriptional variations, and this has in breast cancer led to a subclassification improving breast cancer taxonomy [29].

In the year 2000, Perou et al. identified four breast cancer subtypes based on differences in tumor gene expression patterns [29]. In 2001 and 2003, Sorlie et al. evaluated this classification to include at least five different subgroups [30, 31]. Gene expression profiling using complementary DNA microarrays and hierarchical

clustering separated the tumor samples into two main branches; estrogen receptor (ER) positive and ER negative tumors. ER positive tumors highly express many of the genes expressed in normal luminal epithelial cells, and are further divided into the luminal A and luminal B subgroups. The ER negative tumors are separated into three subgroups; (i) the basal-like group highly expressing genes expressed in normal myoepithelial/basal cells, (ii) the ERBB2 positive group with elevated expression of the *ERBB2* and genes in the proximity, and (iii) the normal breast-like group expressing genes known to be expressed in normal breast tissue like adipose tissue and non-epithelial tissues [31]. Luminal A is the most frequent subtype (~40%) while the ERBB2 positive group occur least frequently (~10%), a distribution observed to be conserved across ethnic groups [32]. In 2007, Herschkowitz et al. identified a new human subtype, claudin-low, characterized by decreased expression levels of genes involved in tight junctions and cell-cell adhesion [33], but this subgroup is currently not commonly used in breast cancer subclassification [32].

The respective subgroups have also been correlated to patient survival, proven a significant difference in clinical outcome. The outcome, overall and relapse-free survival, is most beneficial in the luminal A subgroup and least beneficial in the ERBB2 positive and basal-like subgroups (figure 5). *ERBB2* overexpression and inactivation of the *TP53* tumor suppressor gene are poor prognostic factors frequently detected in the ERBB2 positive group. *TP53* inactivation is as well a common event in the basal-like group. The luminal B subgroup also displays poor outcomes, in particular seen in a long follow-up [34]. A recent study by Curtis et al. analyzing gene

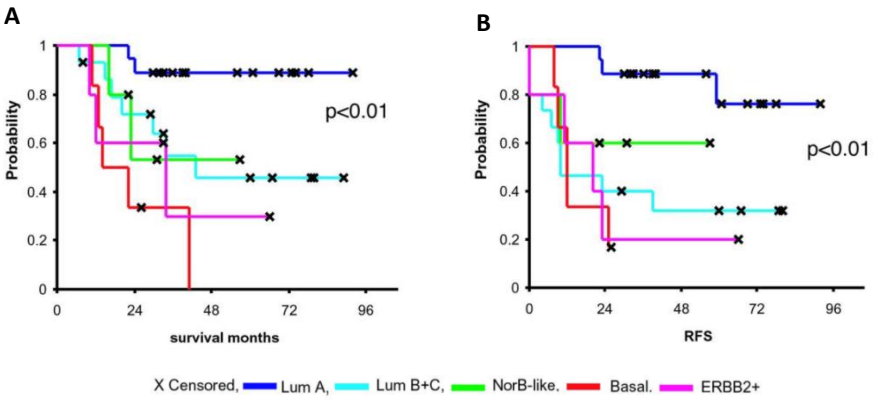


Figure 5: Overall (A) and relapse-free (B) survival analysis of 49 breast cancer patients based on breast cancer subclassification [30].

expression patterns in 2000 primary breast tumors expanded the subclassification system by identifying a total of ten subgroups [35]. The subgroups reflect the complexity in breast cancer biology, indicating that breast cancer might be an assembly of distinct diseases with different therapeutic requirements [30, 31].

1.3. *TP53 – Tumor Protein p53*

The p53 protein was discovered in 1979 as an associated protein to the Simian Virus 40 large T antigen [36, 37]. The p53 association to this viral oncoprotein led to the hypothesis of p53 as a cellular oncogene product with cancer promoting abilities [38]. About ten years passed before the true identity of p53 as a tumor suppressor gene product was revealed [39]. Since then, *TP53* has proven essential in human cancer prevention and been the target of extensive studies, resulting in fame and bynames like *"the Guardian of the Genome"* in the field of cancer genetics [40].

Somatic *TP53* mutations are reported in approximately 50% of all human cancers, although the mutation frequency varies between different cancer types. This makes *TP53* the most frequent altered gene in cancer, reflecting the cancer preventive importance of this tumor suppressor. Inherited *TP53* mutations confirm the tumor suppressive significance as the underlying cause of Li-Fraumeni Syndrome (LFS), a hereditary disorder predisposing to several types of early-onset cancers [41].

1.3.1. Activation of *TP53* and cellular responses

The p53 transcription factor regulates various cellular mechanisms by binding its core domain to specific p53-response elements in its myriad of target genes. p53-activation is a cellular stress response where multiple stimuli like oncogenic activity, hypoxia, nucleotide depletion and DNA damage (figure 6) leads to nuclear p53 accumulation and transcriptional transactivation of p53 target genes [40, 42]. Active wild type (wt) *TP53* promotes tumor suppressive actions by regulating normal cellular growth and survival by inducing cell cycle arrest, DNA repair, replicative senescence and/or apoptosis in stressed or damaged cells to prevent the rise of abnormal clones [42, 43].

The p53 pathway involves hundreds of genes dedicated to the work of genomic integrity maintenance. Intrinsic and extrinsic stress stimuli are the input

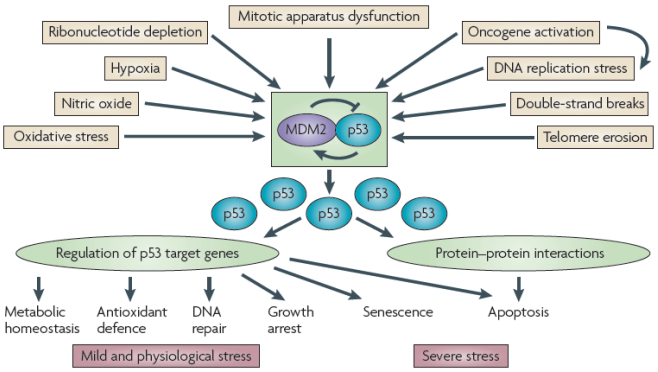


Figure 6: A simplified scheme of the p53 pathway. MDM2 is the main p53 regulator maintaining constantly low p53 levels under normal conditions. Various stress stimuli activate p53 which mediates downstream effects by transactivation of target genes and protein-protein interactions. Severe cellular stress results in irreversible senescence or apoptosis, while milder stress stimuli induce transient cell cycle arrest and repair mechanisms [40].

signals to pathway activation, and upstream p53-mediators sense and transduce these signals to initiate a functional p53 response. In non-stressed cells, p53-levels are almost undetectable due to very short p53 half-life. This is due to p53-inactivation by the main p53-regulator MDM2, an E3 ubiquitin ligase mediating polyubiquitylation and protosomal degradation of p53. MDM2 also inhibits the biochemical activity of p53 by sterical blocking the p53 transactivation domains necessary in transcription initiation. In presence of cellular stress stimuli, the p53-mediators activate and stabilize p53 by post-translational modifications simultaneously as the p53-repressive activity of MDM2 is inhibited. Depending on the initiating stress stimuli, specific target genes are transcriptional transactivated. The synthesized gene products fulfill the cellular p53-mediated responses by inducing a reversible cell cycle arrest if the damage is repairable, or an irreversible senescence or apoptosis in the case of more severe damages [40, 42].

1.3.2. *TP53* mutations in breast cancer

The *TP53* mutation frequency is in breast cancer estimated to 20–40%, indicating that inactivation of *TP53* is an important step in breast cancer development [44]. Mutated *TP53* loses its antiproliferative properties and contributes to tumorgenesis by impaired cell cycle control [40]. In cancers generally, *TP53* mutations are primarily missense mutations located in the DNA binding domain, even though tumor suppressor genes most commonly are inactivated by frameshift or nonsense mutations. In breast cancer, *TP53* mutations occur most frequently in advanced stages and in breast cancer

subtypes of more aggressive behavior, like the ERBB2-positive and basal-like tumors [41].

The discovery of inherited *TP53* mutations as the underlying cause of LFS was a significant clue associating *TP53* to breast cancer development. LFS patients are predisposed to several types of cancer, where breast cancer occurs as the most frequent [41, 44]. *TP53* mutations are believed to be an early event in breast cancer development, since mutations (frequencies ranging from 0–40%) are observed already at DCIS stages of the disease. In addition to advanced stages and aggressive subtypes, younger patients and patients with inherited *BRCA1* and *BRCA2* mutations also show increased somatic *TP53* mutation rates [44]. The association between *TP53* polymorphisms and breast cancer risk has also been investigated. A significant association has been reported in the most common *TP53* SNP, an Arg/Pro polymorphism (rs1042522) in codon 72, exon 4, but the results remain inconclusive [45, 46].

1.3.3. *TP53* as a prognostic and predictive marker in breast cancer

A large number of studies have investigated the association of somatic genetic alterations in *TP53* to the therapeutic prediction and prognosis in breast cancer [44]. A predictive factor is a marker associated with the prediction of treatment-specific responses, while a prognostic factor is a marker associated with clinical outcome, such as overall and disease-free survival, at the time of diagnosis if the disease is left untreated [23].

In 1999, Pharoah et al. published a meta-analysis concerning the prognostic value of *TP53* in breast cancer. 16 studies with over 3500 patients were included, and despite contradictory results, most studies correlated *TP53* mutations to a significantly poorer prognosis compared to wt *TP53* breast cancer cases. *TP53* was in several of these studies also identified as an independent prognostic marker [47]. Comparable results were published by Olivier et al. in 2006, where 18 out of 20 studies identified mutated *TP53* as a marker of poor prognosis in breast cancer [48]. Any inconsistency in these studies may have been caused by a possible publication bias where non-significant findings were left unpublished, or by methodological differences [47].

TP53 mutations have primarily been detected by immunohistochemistry (IHC) and DNA sequencing. Mutated *TP53* often results in an inactive, but stable protein accumulating in the nucleus of the tumor cells. By the use of antibodies, these

mutations are detectable by IHC. Mutations resulting in unstable or truncated p53 proteins are on the other hand not detectable, and nuclear wt p53 accumulation in response to cellular stress may be a source to false positive results using this method. DNA sequencing identifies *TP53* mutations with a greater specificity and sensitivity than IHC. Still, an underestimate probably exist because most studies only sequenced the conserved DNA binding domain (exon 5-8), while nearly 10% of the mutations are found outside this area. DNA sequencing shows a stronger association between *TP53* mutations and breast cancer prognosis than what is attainable with IHC [44, 47]. This could be of clinical value by using *TP53* status in the process towards more individualized treatment of breast cancer patients [48].

The predictive value of *TP53* in breast cancer is less clear than the prognostic implication. The *TP53* antiproliferative abilities of cell cycle control, DNA repair and apoptosis have led to several studies investigating the possibility of *TP53* mutations as therapy response predictors by the use of DNA-damaging agents like chemotherapy and radiation [44]. The results from these studies are conflicting, but indicate that *TP53* mutations are associated to poor treatment responses [44, 49], and imply that a functional p53 pathway is required to induce drug-mediated cell cycle arrest or apoptosis [48]. Publications by Takahashi et al. (2008) and Miller et al. (2005) present *TP53* mutation status alone as a weak predictor, and that the expression profile of a gene set reflecting the *TP53* mutation status signature is a more accurate predictive tool to estimate breast cancer outcome. The Takahashi and Miller publications present two different gene sets (genes significant differentially expressed in mutant and wt *TP53* breast cancer cases) of 33 and 32 genes, respectively. Unexpectedly, the gene sets do not overlap, but both show a stronger association to specific therapy response predictions and clinical outcome than *TP53* status alone [49, 50].

1.4. *WRAP53 – WD repeat containing, antisense to TP53*

1.4.1. Gene nomenclature and discovery

WRAP53 is the abbreviated form of *WD repeat containing, antisense to TP53*, which is the official gene name approved by the HUGO Gene Nomenclature Committee. The nomenclature is derived from the antisense and protein coding characteristics of the gene [1], and the gene products structural homology to members in the WD40

protein family [2]. The *WRAP53* characteristics were discovered by three different research groups, and the findings were published almost concurrently in the beginning of 2009.

WRAP53 as an antisense gene to *TP53* and a p53-regulatory transcript was discovered by Mahmoudi and colleagues at Cancer Centrum Karolinska (CCK), Karolinska Institutet [1]. Venteicher and colleagues at Stanford University identified the *WRAP53* protein as a telomerase subunit, and proved at the same time its involvement in telomerase trafficking to the Cajal bodies[2] (CBs) and telomere synthesis [2]. Tycowski and colleagues at Yale University followed by identifying the *WRAP53* protein as a protein promoting transport of a specific subgroup of small nuclear RNAs (snRNAs) to the CBs [51]. Since the gene discoveries, several different gene names have been in use *(TCAB1 / WDR79 / FLJ10385 / DKCB3*), but the gene will in this thesis be referred to as *WRAP53*.

1.4.2. *WRAP53* – a cis-antisense transcript to *TP53*

WRAP53 is located on the short arm of chromosome 17 (17p13.1), on the opposite DNA strand and upstream of *TP53*. Mahmoudi and colleagues at CCK cloned the *WRAP53* gene, and identified three different non-coding start exons and at least 17 alternatively spliced transcript variants (figure 7). The start exons were named 1α, 1β and 1γ, each contributing to different gene isoforms, which from here will be referred to as *WRAP53α*, *WRAP53β* and *WRAP53γ* [1].

The chromosomal location of *WRAP53* results in a direct overlap between exon 1α in the *WRAP53α* isoform and the first exon in *TP53* [1, 52]. The genes overlap in a head-to-head fashion (5' to 5') by up to 227 base pairs (bp), depending on the *WRAP53* and *TP53* transcription start sites that might vary between different transcripts (figure 8). The overlap between the *WRAP53γ* isoform and *TP53* will only give overlap between precursor mRNAs, not the mature transcripts. The *WRAP53β* transcription start site (i.e. exon 1β) is located downstream of *WRAP53* exon 1α and lack p53 complementarity. The *TP53* overlapping parts of *WRAP53α* will on the other hand give transcripts with perfect complementarity, somewhat makes *WRAP53* a natural antisense transcript (NAT) of *TP53* [1].

[2] CBs are membraneless organelles found in the nucleus of plant and animal cells involved in several functions like RNA modifications, assembly and maturation of ribonucleoproteins (RNPs) and telomere synthesis [61].

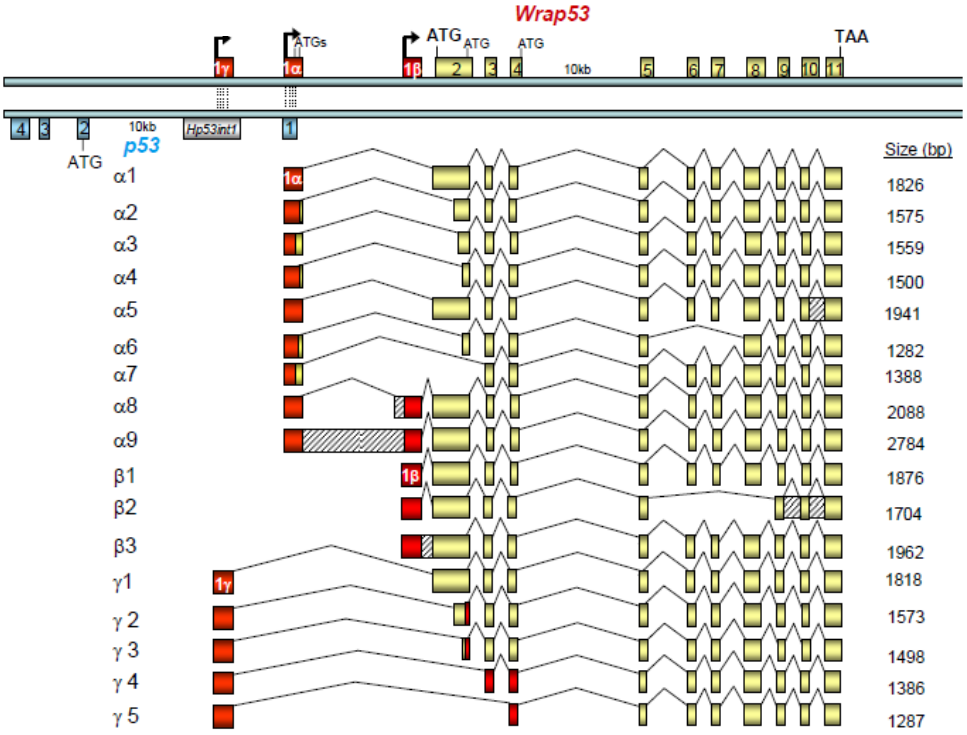


Figure 7: *WRAP53* splice variants.
Yellow bars equal *WRAP53* translated regions while red bars equal untranslated regions. Start (ATG) and stop (TAA) codons for full length WRAP53 protein is written in bold, while alternative start codons are indicated as well. Dash lined bars equal introns included in the mature transcript [53].

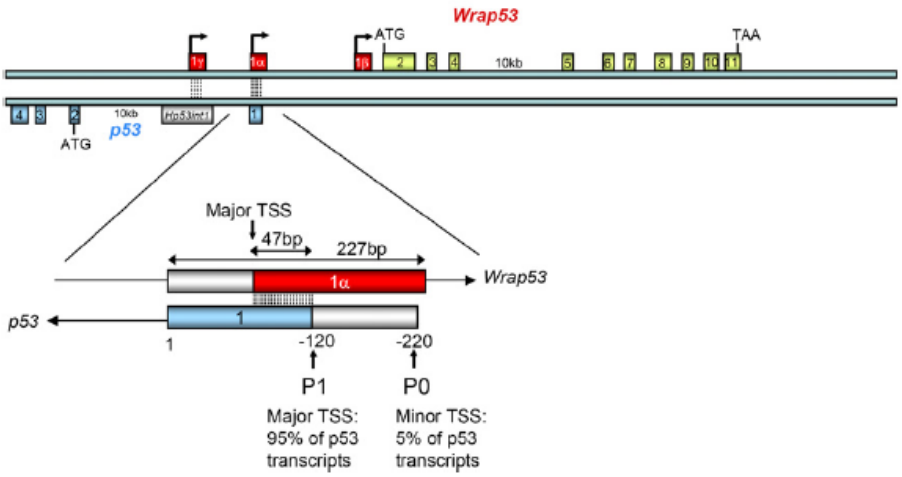


Figure 8: Organization of the *WRAP53* gene and the overlapping regions of *WRAP53* and *TP53*.
Yellow bars equal *WRAP53* translated regions while red bars equal untranslated regions. The start (ATG) and stop (TAA) codon indicate the *WRAP53* coding region. The *WRAP53α* major transcription start site is indicated in the figure, but additional upstream sites exist. *WRAP53β* and *WRAP53γ* are initiated from separate transcription sites [1].

NATs are a group of regulatory single stranded RNAs generated through antisense transcription, a mechanism promoting transcription from the opposite strand

of a protein-coding sense strand. Bidirectional transcription like this makes NATs display transcript complementarity to their corresponding sense mRNA transcripts, affecting mRNA stability, transport and/or translation by transcriptional and post-transcriptional mechanisms [54, 55]. NATs were first described as gene regulatory elements in bacteria, but antisense transcription is now also known as a widespread phenomenon in eukaryote genomes, influencing eukaryote gene expression [56].

Differences in functional properties categorizes NATs into two main groups; (i) *trans*-encoded antisense RNAs transcribed from separate, non-overlapping loci sharing complementary sequences, and (ii) *cis*-encoded antisense RNAs transcribed from overlapping loci on opposite DNA-strands [55]. Hybridization of sense and antisense transcripts generates sense-antisense pairs (SAPs), but SAPs formed by *trans*-encoded NATs will because of the non-overlapping loci just show partial complementarity due to interruption by multiple mismatching base pairs. This give *trans*-encoded NATs the ability to base pair with and regulate the expression of several different sense transcripts, somewhat is the case in microRNA regulation. On the other hand, *cis*-encoded NATs base pairs perfectly with their sense transcripts, resulting in unique relationships to their sense counterparts [56]. *WRAP53* belongs to the group of *cis*-encoded NATs, regulating the expression of its overlapping gene *TP53* [1].

The presence of NATs has been known since their discovery in 1981, but their functional significance has remained obscure. Antisense regulation is largely presumed to depend on complementary base pairing, but a model presented by Munroe and Zhu in 2006 propose five different (Class I-V) antisense regulation mechanisms. Class I, II and III are transcriptional regulation models where the sense gene expression may be influenced by (i) the competition or sharing of transcriptional factors by overlapping genes, resulting in negative correlation or co-expression respectively (Class I), (ii) by bidirectional transcriptional interference due to DNA molecule constraints like RNA polymerase collisions (Class II), or (iii) by SAPs recruitment of transcriptional promoting or inhibiting factors generating epigenetic alterations like DNA-methylations or chromatin remodeling (Class III). The two post-transcriptional regulatory mechanisms are like Class III carried out by the formation of SAPs. In Class IV, RNA duplex formation may mask sense-specific binding sites required for expression, preventing the binding of expression promoting factors. In Class V, SAPs recruits factors that may influence downstream gene expression. The

production of short interfering RNAs (siRNA) from the double stranded SAPs is one example, resulting in degradation of corresponding mRNA transcripts and translation inhibition [55].

Even though SAPs are believed to be of great impact in antisense regulation, detection of endogenous RNA duplexes in human cells has shown itself difficult. This is also the case for WRAP53/p53 duplexes, somewhat indicates that RNA duplexes might be transient and/or labile cellular regulatory mechanisms [1].

1.4.3. Antisense *WRAP53*-mediated regulation of p53

Antisense *WRAP53* transcription gives rise to RNA transcripts which post-transcriptionally regulates the cellular levels of p53. The regulatory effect is restricted to the *WRAP53α* isoform, and is achieved by hybridization of the antisense exon 1α in WRAP53α mRNA to the sense exon 1 localized in the 5' untranslated region of p53 mRNA (figure 8). Antisense transcripts are potential regulators of their corresponding sense transcripts stability, transport and/or translation, and the antisense *WRAP53*-mediated regulation mechanism has shown to be of importance in the cell protective p53-responses generated upon exposure to stressful stimuli [1, 52].

Gene expression studies of normal human tissues and cancer cell lines further revealed a positively correlated *WRAP53α* and *TP53* expression, although the expression of *TP53* occur at 100-fold higher levels than *WRAP53α*. Further cell line studies showed that *WRAP53α* knockdown by siRNA transfection decreased cellular p53 RNA levels by 83% and suppressed induction of the p53 protein by DNA damage, while *WRAP53α* overexpression increased p53 RNA levels 3-fold. Overexpressed *WRAP53α* also increased the cellular level of p53-induced apoptosis. In the same study, blockage of the WRAP53α/p53 RNA hybridization reduced p53 RNA levels similar to *WRAP53α* knockdown, indicating that the *WRAP53*-mediated p53 regulation is sustained through this RNA-RNA interaction. The *WRAP53*-regulation of p53 occurs in a non-reciprocal manner, meaning that altered expression levels of *TP53* do not influence *WRAP53* expression [1].

The results from these studies strongly indicate that *WRAP53* plays a critical role in the regulation of p53. It may seem that antisense *WRAP53α* transcripts are essential in maintaining basal levels of p53, but also in the induction of p53-responses by stabilizing the p53 mRNA and preventing its degradation upon DNA damage. *WRAP53α* is also able to regulate levels of mutated p53, rising the interesting

question concerning *WRAP53* as a new molecular target in treatment of *TP53* mutated human cancers [1, 52].

1.4.4. Discovery of the WRAP53 protein (TCAB1)

In contrast to most regulatory RNAs, *WRAP53* encodes a protein. The WRAP53 protein consists of 548 amino acids and carries six WD40 repeat domains [57] which make it homologous to proteins in the WD40 protein family. WRAP53 was recently identified as a subunit in the telomerase complex, where it is essential in telomerase trafficking to CBs and in telomere synthesis [2, 58]. An alternative protein name used is Telomerase Cajal Body protein 1 (TCAB1), which is descriptive of the telomerase-associated protein function [57].

Telomerase is a RNP complex catalyzing elongation of the chromosome ends by the addition of telomeric repeats (TTAGGG) [19]. Telomerase activity is almost absent in most somatic tissues, but is detected in germline cells and in early embryonic development to obtain telomere lengths sufficient for life [19, 59]. Telomeres protects the genome from losing genetic information during DNA replicative chromosome shortening, which is an important key factor in maintaining cellular genomic stability. Absence of telomerase activity leads to progressive telomere shortening. Critically short telomeres are recognized as double strand DNA breaks, and may lead to telomere fusions and cellular replicative senescence or apoptosis [59, 60]. Telomere shortening is a tumor suppressive action, reflected in the fact that about 90% of human tumors reactivate cellular telomerase activity [18].

The telomerase holoenzyme complex consists of at least three main components; (i) the enzymatic subunit telomerase reverse transcriptase (TERT), (ii) the telomerase RNA component (TERC) which is the RNA template, and (iii) dyskerin, a RNA-binding protein involved in assembly and stability of telomerase and other cellular RNPs [58]. The telomerase complex also includes several associated proteins necessary to proper *in vivo* function [60]. The TERC subunit is a non-coding snRNA, i.e. small nuclear RNA molecules found in eukaryotic organisms guiding the modification of other RNA molecules [61]. Small nucleolar RNAs (snoRNAs) is a large subgroup of the snRNAs involved in ribosomal RNA modifications in the nucleoli[3] and splicosomal RNA modifications in the CBs [2, 51]. The snoRNAs are

[3] The nucleoli are membraneless subnuclear organelles involved in ribosomal RNA synthesis, processing and assembly [19].

further divided into two groups based on the presence of one out of two different conserved structural motifs; (i) the C/D box snoRNAs which guide methylation, and (ii) the H/ACA box snoRNAs which guide pseudouridylation [51]. Dyskerin is able to recognize the H/ACA box motif, and is essential in assembly and stability of snoRNPs, telomerase included. The mechanism behind the selective transportation of snoRNPs to the nucleoli or CBs was for a long time poorly understood, but a possible explanation was given in a publication by Venteicher et al. in 2009, where the WRAP53 protein was identified as a subunit of the telomerase holoenzyme complex [2].

WRAP53 was discovered by the identification of possible new dyskerin-interacting proteins. In addition, the study showed that WRAP53 was able to interact with TERT and TERC, and that telomerase activity depended on the presence of WRAP53. These observations indicated that WRAP53 was a part of the enzymatic active human telomerase complex. Further studies showed that WRAP53 accumulated in the CBs, and that it specifically interacted with a subgroup of snoRNAs that are called small Cajal body specific RNAs (scaRNAs) [2]. The scaRNPs are similar to the snoRNPs, but more complex because they contains two structural motifs; a combination of one C/D box and one H/ACA box, two C/D box motifs or two H/ACA box motifs. scaRNAs that contains one or two H/ACA box motifs also contain another structural motif called the CAB box. The CAB box is a conserved tetranucleotide (UGAG) present in the H/ACA motif, essential for CB localization of scaRNAs [51]. Specific binding of WRAP53 to this CAB box is what regulates the specific localization to and retention of scaRNAs, including TERC, in the CBs [2, 51].

1.4.5. The WRAP53 protein and telomerase activity

Identification of WRAP53 revealed a protein structural homologous to proteins in the WD40 repeat protein family [2]. The WD40 repeat proteins are a large protein family abundantly expressed in eukaryotes. The name is derived from the structure of the repeated domain which contains a conserved tryptophan (W) and aspartic acid (D) dipeptide and expands about 40 amino acids in length. WD40 repeat proteins are involved in a diverse range of cellular functions such as transcriptional regulation, cell cycle control, apoptosis and signal transduction, but the most common function is their implication in protein complex assembly [62].

The discovery of WRAP53 as a telomerase subunit also revealed that the protein is of importance in telomerase trafficking and telomere synthesis. Telomerase activity depends on WRAP53-mediated trafficking of TERC to the CBs, and further trafficking to the telomeres during the cell cycle S-phase for telomere elongation. This specific trafficking is mediated by the interaction between WRAP53 and the TERC CAB box domain. WRAP53 depletion prevents TERC from localizing to the CBs [2], but active complexes containing TERT, TERC and dyskerin can still be created. The effect of WRAP53 depletion is observed as reduced amounts of TERC localized in the CB and by the telomeres. This indicates that reduced WRAP53 levels repress telomerase functionality, and that WRAP53-mediated CB localization is essential to telomere maintenance [2, 58]. Discoveries by Mahmoudi and colleagues show that the WRAP53 protein mainly is generated from the *WRAP53β* isoforms [3], and that the protein is essential in CB formation. *WRAP53* knockdown makes existing CBs collapse, and prevents new formation. CBs are not vital cellular organelles, but their absence probably reduce the efficiency of CB-mediated processes [63].

CAB box mutations are other elements influencing CB-specific trafficking of scaRNAs, including TERC. CAB box mutated scaRNAs are not able to bind to WRAP53, and mislocalize to the nucleoli [51]. In the same way as WRAP53 depletion, CAB box mutations inhibit telomere synthesis and leads to progressive telomere shortening [2, 51]. The fact that telomerase activity distinguishes cancer cells from normal somatic cells, has for a long time kept telomerase as a promising therapeutic target in cancer therapy. Unfortunately, the development of drugs targeting the enzymatic core of telomerase has been difficult. The discovery of WRAP53 may now open to new therapeutic strategies, as dysfunctional WRAP53 results in telomere shortening [58].

1.4.6. *WRAP53* in cancer

Several studies imply that *WRAP53* may be involved in tumor promoting actions because of its impact on p53-regulation and telomerase activity, both cellular functions altered in tumorgenesis.

Discoveries by Mahmoudi and colleagues indicate that *WRAP53* might have oncogenic properties and be involved in cellular transformation. This statement is based on the observation of elevated *WRAP53* expression levels in cancer cells compared to immortalized and normal cells. Overexpressed *WRAP53* promoted

anchorage-independent colony growth, which is a cancer cell characteristic. *WRAP53* knockdown in cancer cells by siRNA treatment induced massive apoptosis through the mitochondrial-mediated pathway, but the same did not occur in non-transformed cells. This imply that cancer cells are more sensitive to depleted *WRAP53* levels, and that cancer cells depend on *WRAP53* expression to survive, somewhat strengthens the *WRAP53* oncogene hypothesis [3]. The same study also generated results indicating that *WRAP53* may be of prognostic value in primary head and neck cancer. *WRAP53* overexpression in cell lines obtained from primary head and neck tumors correlated to poor patient outcome, while lower expression levels correlated to a more beneficial outcome. This was further confirmed by the observation that increased *WRAP53* expression levels associated to decreased sensitivity to radiation therapy, a common therapeutic strategy used in head and neck cancer [3].

Dysfunctional WRAP53 protein disrupts telomerase trafficking and is one of several underlying causes of Dyskeratosis congenita (DC). DC is an inherited syndrome where telomere shortening and reduced stem and progenitor cell function results in defective tissue maintenance, bone marrow failure and cancer predisposition. Zhoung et al. identified four different missense *WRAP53* mutations located in various exons in two out of nine unrelated DC patients. Absence of these genetic alterations in control subjects declined the findings as common polymorphisms. Compared to wt *WRAP53* patients, reduced *WRAP53* expression and impaired CB-specific accumulation were observed in the mutated patients, indicating mutation-mediated protein deficiency and possible novel cancer susceptibility factors in DC patients [64].

Specific genetic alterations in *WRAP53* are also related to increased cancer risks. Studies show that two linkage disequilibrium[4] (LD) SNPs, an Arg/Gly polymorphism in codon 68 (rs2287499), exon 2, and a Phe/Phe polymorphism in codon 150 (rs2287498), exon 3, is significantly associated with an increased risk of ER negative breast cancers [65]. The rs2287498 polymorphism is as well associated with an increased risk of invasive ovarian cancer [66].

[4] LD is a non-random association of alleles at separated but linked loci. LD loci lie so close that they are not segregated by recombinations and because of this inherited together as a package [19].

1.5. Aims of the study

The *WRAP53* gene was identified as an antisense transcript to *TP53* and its regulatory effect has proven essential in the initiation of p53 pathway responses. *WRAP53* depletion is discovered as a pro-apoptotic factor in cancerous cells, and its gene product as a subunit in the telomerase holoenzyme complex. Despite limited knowledge about *WRAP53* functionality, the gene is proposed to be involved in carcinogenesis. The p53 pathway and telomerase activity are cellular mechanisms frequently altered in cancer cells by inactivation and upregulation respectively, and *WRAP53* expression seems to be necessary both to p53 and telomerase function [1-3]. Impairment in such cellular key mechanisms by genetic alterations in *WRAP53* might be a contributor to cancer development. The overall aim in this thesis is to study the importance of and increase the knowledge about *WRAP53* in breast cancer.

The intermediate aims in this study are;
i) To perform a *WRAP53* mutation analysis on DNA from primary breast carcinomas searching for somatic genetic alterations that might be associated with development and progression of breast cancer. Such findings may contribute to the understanding of a new predictive and/or prognostic marker in breast cancer. Prior to the mutation analysis a new DNA sequencing method was introduced as a sub-aim of this study.

ii) To perform gene expression studies in *WRAP53* depleted breast cancer cell lines. This functional study may increase the knowledge about genes regulated by *WRAP53*, directly or indirectly. Such knowledge may contribute to better understanding of the cellular signaling pathways and networks *WRAP53* is involved in, and potentially to the identification of novel therapeutic targets in breast cancer therapy.

2. Materials

2.1. Ethical considerations

Medical research projects using human biological samples and personal clinical data are imposed to follow valid ethical guidelines. In the current study, this is covered by approval of the Regional Committees for Medical Research Ethics and by written informed consent obtained from the patients prior to the study. All samples and clinical data were de-identified, and national and institutional guidelines considering biobanking are followed.

2.2. Patient Materials

The patient materials used in this study is a series of 212 primary breast cancer cases sequentially collected at Ullevål University Hospital from 1990–1994 (from here referred to as ULL-samples). The mean age of the patients was 64,4 years (ranging from 28,2–91,5 years), and all patients were treated in accordance with Norwegian national guidelines at the time of diagnosis [34]. The Ullevål cohort also contains blood samples from 119 of the patients collected from 1994–1996. Time range from diagnosis to blood collection varied from 0–6 years [65]. Clinical patient information was last updated in 2006, resulting in an observation time from 12–16 years [34].

Tissue from the primary breast tumors were snap frozen and stored at –80°C. DNA was isolated from both tumor tissue and peripheral leukocytes using a standardized method of phenol-chloroform extraction and ethanol precipitation (Model 340A Nucleic Acid Extractor, Applied Biosystems) [34].

In this study, genomic DNA from 175 ULL tumor samples was included in the *WRAP53* mutation analysis to search for genetic alterations that might be associated to breast cancer. DNA from blood samples available were used to distinguish between the findings of germline and somatic origin. Sample excluding criterions were inadequate sample volumes and low DNA concentrations (< 5 ng/µL DNA quantified by the NanoDrop® ND-1000 Spectrophotometer, Saveen Werner).

2.3. Cell lines

Breast cancer cell lines are extensively used to investigate breast cancer biology, pathology and therapy responses [67]. Among all commercial cell lines available, MCF-7 and MDA-MB-231 are the most commonly used in breast cancer research [68]. These were also the cell lines used in the *WRAP53* knockdown and gene expression study, chosen because they represent two main types of breast cancer.

MCF-7 is a human breast adenocarcinoma cell line derived by the Michigan Cancer Foundation (from where the name is derived) in 1973 [68]. The cell line is established from a pleural effusion obtained from a 69 years old female Caucasian. MCF-7 has preserved several differentiated mammary epithelium characteristics, and is because of this useful in breast cancer studies [69]. The MCF-7 cell line is ER positive, progesterone receptor (PR) positive, HER2 negative, wt *TP53* and subtyped as luminal [67]. The MCF-7 cell line was obtained from Interlab Cell Line Collection, Genova, Italy.

MDA-MB-231 is a human breast adenocarcinoma cell line derived by the M. D. Andersons Hospital and Tumor Institute, Texas, in 1973 [70]. The cell line is established from a pleural effusion obtained from a 51 years old female Caucasian [70, 71]. MDA-MB-231 is ER negative, PR negative and HER2 negative (also known as triple negative breast cancer). It harbors a *TP53* mutation (Arg>Lys in codon 280, exon 8 [72]) and is subtyped as basal-like [67]. The MDA-MB-231 cell line was obtained from American Type Culture Collection, Manassas, Virginia, USA.

3. Methods

3.1. Patient sample preparation

The *WRAP53* mutation analysis was performed on genomic DNA isolated from the sample series collected at Ullevål University Hospital from 1990–1996 (chapter 2.2).

3.1.1. DNA isolation

DNA was previously isolated from the ULL tumor (ULL-T) and blood (ULL-B) samples by a phenol-chloroform extraction method followed by an ethanol precipitation using the Model 340A Nucleic Acid Extractor (Applied Biosystems) [34]. Phenol-chloroform extraction is a liquid-liquid extraction method based on sample component separation between an aqueous and an organic phase, and the method is commonly used to purify nucleic acids from complex biological samples [73].

To isolate the DNA, samples were mixed with a preheated solution of Lysis Buffer and Proteinase K to heat-inactivate endogenous nucleases. The Lysis Buffer contained a chaotropic agent (urea) and an anionic surfactant (n-lauroyl sarcosine), lysing the cells and denaturizing nucleases and proteins. The proteins were further degraded by Proteinase K, and a chelating buffer agent, cyclohexaminediamine tetraacetic acid, restrained DNase activity by binding the enzymatic cofactors of divalent cations.

The separation phase was induced by the addition of phenol-chloroform reagent (50/50 v/v) to the lysate. Nucleic acids were restrained in the upper aqueous phase, while peptides were extracted to the lower organic phase and disposed to waste. Depending on the protein concentration in the sample, the phenol-chloroform step could be repeated to increase the nucleic acid purity.

Addition of 95% ethanol to the sample gave precipitation of the DNA. By pre-adding sodium acetate, sodium bound to the DNA phosphate groups and eased the DNA precipitation. Precipitated DNA was transferred to a sample tube and dissolved in Trizma® base/Ethylenediaminetetra-acetic acid (EDTA)-buffer. The chelating abilities of EDTA inhibited DNase activity by binding divalent cations [74]. The isolated DNA was stored at 4°C.

3.1.2. DNA quantification

The DNA concentrations of the patient samples analyzed in this study was quantified using a NanoDrop® ND-1000 Spectrophotometer (protocol in Appendix A). NanoDrop® ND-1000 is a cuvette free spectrophotometer preforming absorbance measurements in the 220–750 nm spectrum using a sample volume of 1,0 µl. Two fiber optic cables in contact with the sample and a xenon lamp light source accomplish the sample measurement, and the transmitted light intensity is detected. The NanoDrop® ND-1000 Spectrophotometer accurately measures the concentration of double stranded DNA (dsDNA) up to 3700 ng/µl without dilution [75].

Nucleic acids absorb electromagnetic radiation at the wavelength of 260 nm, and therefore the absorbance measurements do not distinguish between the different types of nucleic acids (RNA, single stranded DNA or dsDNA), or other compounds that absorb the same wavelength. DNA purity is determined using a ratio of sample absorbance measured at 260 and 280 nm (260/280 ratio). A ratio ~1,8 indicates pure DNA solution, while ratios < 1,8 indicate presence of contaminants like proteins, phenols or other compounds that absorb 280 nm [75]. High DNA purity is not a requirement in the polymerase chain reaction (PCR), but the presence of contaminants may inhibit the DNA polymerase and influence the PCR efficiency [19].

The measured DNA concentrations were used in calculations to make PCR sample dilutions containing 5 ng/µl DNA. The samples were diluted in DNase/RNase free water (GIBCO, Ref 10977-35).

3.2. Introducing the BigDye® Direct Cycle Sequencing Kit Method

Prior to the *WRAP53* mutation analysis (described beneath) a new DNA sequencing method was introduced at the Department of Genetics, and the testing and adjustments of the method was a subtask of the work presented in this thesis. The existing method used the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems), but an improved version, the BigDye® Direct Cycle Sequencing Kit (Applied Biosystems) was recently launched in the market. The BigDye® Direct Cycle Sequencing Kit promotes a faster workflow with fewer steps and read lengths starting as close as one base from the primer. Both kits use universal M13 sequencing primers, requiring M13-modified PCR primers in the PCR amplification step.

The workflow of both methods involves DNA template preparation, PCR, agarose gel electrophoresis (optional), PCR product purification, cycle sequencing, sequencing product purification and capillary electrophoresis. The main differences are the steps of PCR and sequencing product purification. The BigDye® Terminator v1.1 Cycle Sequencing Kit requires separate steps of vacuum based filtration to purify the PCR products, and a Sephadex™ gel filtration method to purify the dye terminator sequencing products prior to electrophoresis [76]. In the BigDye® Direct Cycle Sequencing Kit, the PCR product purification step is combined with the cycle sequencing reaction. The BigDye® Direct Sequencing Master Mix contains reagents to perform both cycle sequencing and PCR product purification. Unincorporated PCR primers are degraded, while the M13 sequencing primers are protected from this treatment, resulting in optimal cycle sequencing conditions. The purification of the dye terminator sequencing products are accomplished by adding a bead-based purification solution directly to the samples using the BigDye® XTerminator™ Purification Kit (Applied Biosystems) [77].

The DNA sequencing methods procedures and sequence quality was compared sequencing *TP53*, a gene frequently analyzed at the Department of Genetics, giving a proper basis of comparison. Reduced work load, improved sequence quality and acceptable cost levels were the foundation for introducing the BigDye® Direct Cycle Sequencing Kit as the new standard sequencing method at the Department.

3.3. *WRAP53* mutation analysis

The *WRAP53* mutation analysis was performed using the Sanger sequencing method. All ten coding exons and start exon 1β was sequenced in the search for genetic alterations. The reason for not sequencing the other two *WRAP53* start exons, 1α and 1γ, is because β-transcripts occur more abundantly in cells than α- and γ-transcripts, and that *WRAP53* proteins primarily are synthesized from the β-isoforms [78]. Genetic alterations in the β-transcripts might because of this be of greater importance, especially when it comes to protein function, than the two other transcript isoforms. The DNA sequencing was performed using the Applied Biosystems 3730 DNA Analyzer in accordance to the supplier's instrument protocol [79]. All the exons were

sequenced in both directions based on independent PCR amplifications generated prior to the forward and reverse cycle sequencing reactions, in order to obtain high-quality sequencing data.

The BigDye® Direct Cycle Sequencing Kit procedure (protocol in Appendix B) consists of six main steps. The first step is a PCR amplification followed by an agarose gel electrophoresis as a PCR quality control. The Sanger sequencing reaction generates the sequencing products which are further purified prior to capillary electrophoresis. The last step is SecScape software (v2.7, Applied Biosystems) data analysis and interpretation.

3.3.1. Polymerase chain reaction

PCR is an *in vitro* DNA cloning method that selectively amplifies target DNA sequences by the use of specific primers. In a cyclic process of DNA denaturation, primer annealing and extension, the selected DNA sequence amplifies exponentially. PCR is an extremely robust, but also very sensitive method [19], which highlights the importance of including negative controls without target sequences to monitor the possible occurrence of interfering contaminants.

Altogether, thirteen *WRAP53* primer pairs (Eurogentec) were used to amplify the ten *WRAP53* coding exons and start exon 1β using PCR. The *WRAP53*-specific localization of the primers were controlled before they were designed as described in the supplemental material of the article *"Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita"* (Zhoung et al. 2011) [64] (table 1). The primers were modified with a universal M13 sequence at the 5' end;
M13 forward primer sequence: 5' -TGTAAAACGACGGCCAGT-3'
M13 reverse primer sequence: 5' -CAGGAAACAGCTATGACC-3'

The *WRAP53* primers specificity was tested prior to the mutation analysis by PCR amplification and agarose gel electrophoretic separation of the *WRAP53* exons of current interest. The test was performed on MDA-MB-231 isolated genomic DNA using the BigDye® Direct Cycle Sequencing Kit (Applied Biosystems).

Table 1: *WRAP53* primers used in the PCR amplification step of the mutation analysis

Exon	Forward primer (5')	T_m (°C)	Reverse primer (3')	T_m (°C)	PCR fragment lenght (bp)
2A (exon 1β)	GGGAACGGGAAACCTTCTAA	62,6	GACAGCAGTCCGGAGCTAAC	64,8	371
2B	CTAATCTCCGCTGTGCTTCC	63,7	TCTTCTGCAGGAAGGCTTGT	62,6	350
2C	GGGACCCAGTTTCTCTCTCC	64,8	CTGGAGAAGTGGGTCTCAGG	64,8	311
3	GTGGAGTCTGGGGAGATGAA	63,7	GGGCATCCCTCTCCTAGAAA	63,7	304
4	CAGCCCTAGCCCTACACTTG	64,8	TGCTGCCACAAGAAATTCAC	61,6	414
5	TCTGAGCTCACCCTTGAACA	62,6	CTGACCAGCCCCTCTGATAA	63,7	357
6	ACACCCAGCCTCATTTTTGT	61,6	GGAAGGAAAGGGCTGAAAAC	62,6	392
7	TCATATCTGGGACGCATTCA	61,6	GTACAGAGGACGGCGTGAAC	64,8	411
8A	GCTTGTGACAGACAGCATGG	63,7	TCTCAGGGTGTGACCCCTAC	64,8	363
8B	TCTGTATGCCTGGGATGATG	62,6	ATTGGTGGTCACCTCTCGAC	63,7	383
9	CTGAAGGAGTGCCTGGAGAC	64,8	ACCCTACAGCTGGGCTCTG	64,9	259
10	CCTCTGCCAGCAAATCTCTC	63,7	TCTCTGTGGGCTCAGGAAAC	63,7	351
11	AGAGGGAGCAAGTGTCCTCA	63,7	GCCTGGTTTCAGGACCAATA	62,6	436

3.3.2. Agarose gel electrophoresis

The PCR specificity and PCR product quality was qualitatively controlled by agarose gel (1,5% agarose, recipe in Appendix F) electrophoresis. Electrophoresis is a method separating charged molecules in an electric field, forcing the negatively charged DNA molecules to migrate towards the positive electrode. Because of the negatively charged phosphate groups, DNA molecules have the same netto charge per unit length, making DNA fragment size the factor of separation. In a polymerized agarose gel, small molecules migrates faster and longer than the larger ones, and the use of a DNA-ladder (φX 174-Hae III digest, TaKaRa) makes it possible to estimate the fragment sizes [19].

The results from the electrophoresis were visualized by GelRed™ nucleic acid staining (GelRed™ Nucleic Acid Gel Stain, Biotium) and ultraviolet (UV) gel irradiation. GelRed™ is a fluorescent dye binding nucleic acids through intercalations and electrostatic interactions. UV irradiation excites the dye molecules which subsequently emit the excess energy as visible fluorescent light [80]. The results were developed using the GeneGenius Bio Imaging System (Syngene) and GeneSnap Software (v7.01.07, Syngene).

3.3.3. Sanger sequencing method

DNA sequencing is a term including all methods used to determine the order of the nitrogen bases in the DNA molecule. Sanger sequencing (also known as the dideoxy sequencing or chain termination method) is a frequently used method, and is in many

research communities considered as the golden standard sequencing method. Even though next generation sequencing has occurred as a revolutionizing tool in cancer genome characterization, Sanger sequencing is still commonly used in small-scale experiments. The method is as well used to confirm sequence alterations and complete fragments difficultly sequenced by massively parallel sequencing approaches. Sanger sequencing is based on enzymatic DNA synthesis and random inhibition of the growing chains, creating premature terminated transcript of various lengths. Transcript elongation in DNA synthesis depends on free hydroxyl groups at the deoxynucleotide triphosphate (dNTP) 3' carbon and formation of phosphodiester bonds. The premature termination in Sanger sequencing is induced by the presence of base-specific dideoxynucleotide triphosphates (ddNTPs), analogues to the dNTPs where the 3' carbon hydroxyl groups are replaced by hydrogen atoms. The hydrogen atoms inhibit the formation of phosphodiester bonds and terminates the DNA synthesis [19].

In Sanger sequencing, the ddNTPs terminates the DNA fragment sequenced in every single base position occupied by corresponding dNTPs. Fluorescent labeling of the ddNTPs by dye molecules with different emission spectra makes it possible to determine the base terminating each fragment (figure 9A), and in combination with the fragment lengths determine the order and position of the bases in the DNA molecule [19].

3.3.4. Sequencing product purification: BigDye® XTerminator™ Purification Kit

Purification of sequencing reaction products is a necessary step to ensure high quality capillary electrophoresis sequencing results. The purification step removes excess components like unincorporated fluorescent tagged ddNTPs and salts that might interfere with the results. The BigDye® XTerminator™ Purification Kit (Applied Biosystems) is a bead-based purification method that captures and immobilizes the unwanted components by vortexing the reaction plate. Centrifugation of the vortexed plate sediments the insoluble fraction of beads and captured components, creating a supernatant of purified dye-labeled sequencing products that directly can be injected in the capillary electrophoresis instrument for analysis [81].

The BigDye® XTerminator™ Purification Kit consists of two different solutions; (i) the bead-containing XTerminator™ Solution that captures the excess ddNTPs and salts, and (ii) the SAM™ Solution that enhances the performance of the

XTerminator™ Solution and stabilizes the purified sample. A premix of these solutions is added to the samples after the cycle sequencing reaction, and proper vortexing is important to ensure appropriate mixing and purification [81].

3.3.5. Capillary electrophoresis: Applied Biosystems 3730 DNA Analyzer

The Applied Biosystems 3730 DNA Analyzer is a capillary-based automated DNA sequencer [82]. In capillary electrophoresis, charged molecules are separated inside a narrow capillary tube by the influence of an electric field. To separate DNA fragments of constant size-to-charge ratio, the capillary is filled with a polymer separating the fragments according to size. The samples are electrokinetic injected into the capillary end and migrates with size-dependent velocity towards the positive electrode in the opposite end of the capillary. Near the positive electrode every capillary displays a detection window. The 3730 DNA Analyzer uses a laser-induced fluorescence detection method. The signal from the DNA fragments passing the detection window is captured as a result of laser irradiation generating a ddNTP base-specific fluorescent light. By combining the migration time and the emission spectrum of the fluorescent light, the order of the nitrogen bases in the sequenced DNA fragment are determined (figure 9A) [82, 83].

The Applied Biosystems 3730 DNA Analyzer is a 48 capillary instrument using POP-7™ Performance Optimized Polymer (Applied Biosystems) developed to give longer read lengths and shorter run times. Processing of the raw data file displays the analyzed sample data as an electropherogram (figure 9B) [82].

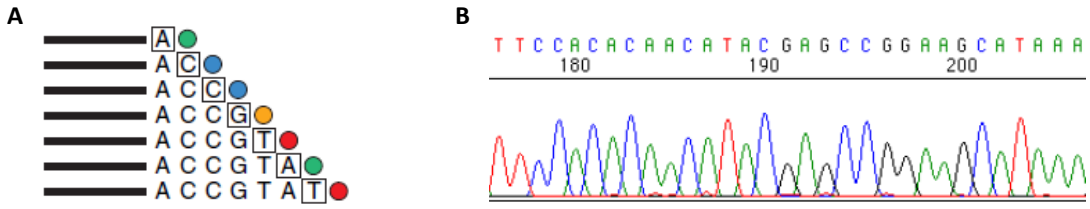


Figure 9: Fluorescent labeled cycle sequencing dye terminator products with squared terminating ddNTPs (A) and sample electropherograms displaying the DNA nitrogen base order by base color-specific peaks (B) (Applied Biosystems, 2009).

3.3.6. SeqScape v2.7 sequencing data analysis

SeqScape v2.7 is an Applied Biosystems software tool used to analyze and identify sample genetic alterations to a consensus sequence. Mutation analysis, SNP discovery

and validation are SeqScape applications [84] used in this study. The sequencing data were analyzed in accordance with the National Center for Biotechnology Information (NCBI) *WRAP53* Reference Sequence NG_028245.1 [85], and comprised start exon 1β, the entire coding *WRAP53* region (exon 2–11) and the intronic sequences of 30 bp prior to and following each exon in the case of splice mutations.

To generate interpretable sample electropherograms, the SeqScape software processes the raw data files obtained from the Applied Biosystems 3730 DNA Analyzer. Some important and necessary processing steps are (i) the multicomponent analysis that filters the four fluorescent dye signals in distinct spectral components to avoid spectral overlap interference, (ii) basecalling that adjusts the fluorescent signals and assigns one base to each peak, and (iii) the mobility shift correction that rectifies changes in the electrophoretic mobility of sequencing products imposed by the association with differently labeled ddNTPs [82]. The forward and reverse sample sequences are assembled and compared to the consensus sequence [85], and deviations from the consensus sequence detected by the software is reported [84].

3.3.7. Statistical and bioinformatics data analyses

The statistical analyses were performed using IBM SPSS version 19. The results from the *WRAP53* mutation analysis were investigated according to potential associations within the detected genetic alterations, and association of genetic alterations in *WRAP53* with breast cancer-specific survival and other known clinical patient parameters. The Chi-square and Fisher`s exact tests were used when appropriate to determine the strength of relation between two categorical variables, like the *WRAP53* genetic alterations and the presence of lymph node metastasis. The Chi-square test is not recommended used if 20% or more of the cells have expected observation counts less than five. The Fisher`s exact test is then the alternative statistical approach [86]. The association between detected *WRAP53* genetic alterations and breast cancer survival was tested using the Kaplan-Meier survival plot and the Logrank test comparing the survival distributions under different conditions. Overall, p-values $\leq$ 0,05 were considered statistically significant. Frequency plots were generated using Microsoft Excel 2010.

The level of genetic linkage in the dataset was investigated using the

Haploview software[5] (v4.2, Broad Institute). Haploview is a bioinformatics tool designed to perform LD and haplotype block analyses and to estimate haplotype frequencies [87]. Haplotypes are loci on the same chromosomal segment that tends to be inherited together as blocks [19]. The level of genetic linkage between loci is measured by statistically calculating the genetic distance between the loci, a measure based on the expected recombination fraction. Genetically linked loci lie close together and do not segregate independently by recombinations, reducing the genetic distance and creates haplotype blocks inherited together for generations [19, 73]. The likelihood of LD were calculated based on a combination of D' and LOD score values. D'=1 is known as complete LD, while D'<1 indicates LD disruption. The LOD score compares the likelihood of obtaining the test data if the loci are linked, versus not linked. A positive LOD score favors linkage, while linkage is not likely present by negative scores [19, 88].

Haploview also checks conformance with the Hardy-Weinberg (HW) equilibrium due to possible genotype selections. The HW equilibrium states that the genetic variation in a population remains constant throughout generations in the absence of disturbing factors, and predicts that genotypes and allele frequencies remains constant because they are in equilibrium [89].

The detected *WRAP53* sequence alterations were investigated using the SNP500Cancer [90] and the NCBI dbSNP [91] databases. The databases contain overview of gene-specific SNPs and other minor genetic variations like small insertions and deletions, and are frequently used databases in SNP studies.

3.4. Gene expression study

Downregulation of *WRAP53* has been suggested to influences the p53 pathway, the telomere synthesis and the level of apoptosis [1-3]. *WRAP53* may also, directly or indirectly, possess other regulatory functions and involvements in cellular pathways, roles completely unknown today. The field of *WRAP53* research aims for a greater understanding of the gene`s function, and in particular knowledge related to cancer diseases. An analysis comparing gene expression patterns in *WRAP53* normal expressing and siRNA-treated *WRAP53* depleted cell lines, in the current study

[5] Available from http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/downloads

focused on breast cancer, is a relevant approach to reach the goal.

The gene expression study was performed in collaboration with Farnebo and colleagues at the Department of Oncology-Pathology, CCK, Karolinska Institutet, Stockholm. The cell culturing, siRNA transfection and RNA isolation was accomplished at CCK, while the gene expression analyses, data processing and interpretation of results were performed as a part of the work included in this study.

3.4.1. Cell culturing

The MCF-7 and MDA-MB-231 cell lines were obtained from the Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, and sent to CCK in a container of dry ice.

The MCF-7 and MDA MB 231 cells were maintained in Dulbecco`s Modified Eagle Medium (HyClone, Thermo Scientific) supported with 10% fetal bovine serum (HyClone, Thermo Scientific) and 2,5 µg/ml placmocine (InvivoGen) at 37°C and 5% CO_2 humidified incubators. The cells were cultured for two weeks before siRNA transfection to ensure that the processes of thawing and new culture establishment not influenced the gene expression patterns of the cells. Cell passaging was performed twice a week to both cell lines [92].

3.4.2. siRNA transfection

The highly reduced *WRAP53* gene expression levels in MCF-7 and MDA-MB-231 was induced using a *WRAP53* siRNA (Qiagen) targeting exon 2 (siWRAP53#2, 5'-AACGGGAGCCTTTCTGAAGAA-3') resulting in *WRAP53* knockdown independent of the different gene isoforms. siRNAs are short double stranded RNAs of 20–25 base pairs in length frequently used in *in vitro* gene silencing [19]. The transient siRNA transfection was accomplished using the HIPerFECT Transfection Reagent (Qiagen) [92], a mix of cationic and neutral lipids that promotes siRNA uptake and intracellular siRNA release [93].

The experimental setup included non-treated control (NTC) cells, siRNA control (siC) transfected cells, and siWRAP53#2 transfected cells for each cell line at two different time points (table 2). The siC used was a negative control siRNA (20 nmol, Qiagen) consisting of lipids and scrambled siRNA oligos to ensure that the transfection procedure itself does not influence the gene expression patterns in the cells. Prior to the transfection, 30000 cells/ml (2 ml/well) was seeded in 6 well plates.

By 24 hours after the seeding, the cells were siRNA transfected using 10 nM siC or 10 nM siWRAP53#2 combined with 6,0 µl HIPerFECT per cell culture [92]. RNA was harvested from single cultures (NTC cells) and biological triplicates (siC and siWRAP53#2 transfected cells, different cultures) 40 and 72 hours after transfection in both cell lines.

Table 2: The experimental setup in the gene expression study. RNA was harvested from single cultures (NTC cells) and biological triplicates (siC and siWRAP53#2 transfected cells) 40 and 72 hours after transfection.

Cell line	Treatment	Replicate
MCF-7	NTC	1
	siC	1
		2
		3
	siWRAP53#2	1
		2
		3
MDA-MB-231	NTC	1
	siC	1
		2
		3
	siWRAP53#2	1
		2
		3

Lipid-mediated transfection is a method where liposomes, lipid-formed vesicles, transfer and release the siRNA to the intracellular environment. Cationic lipids, amphiphilic molecules with a cationic hydrophilic headgroup and a hydrophobic tail, are most commonly used. In aqueous environments, cationic lipids form vesicles with the hydrophilic headgroups on the exterior and the hydrophobic tails in the interior. Complexion of liposomes and siRNAs creates lipoplexes able to interact with the cell membranes and promote endosomal uptake. Small lipoplexes establish an electrostatic interaction between the positively charged lipoplex and the negatively charged cell membrane, while the larger lipoplexes are endocytosed via clathrin coated pits. To avoid endosomal-lysosomal fusion and siRNA destruction, the cationic lipids are often combined with neutral helper-lipids, like cholesterol, which destabilizes the endosomal membrane and mediates siRNA escape. Intracellular siRNAs are processed by the RNA interference pathway and promotes gene knockdown by mRNA degradation and transcription repression [19, 94].

3.4.3. RNA isolation and microarray sample preparation

Total RNA was extracted from the MCF-7 and MDA-MB-231 cell lines using the TRIzol® reagent (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The TRIzol® RNA isolation method is a guanidine isothiocyanate-phenol-chloroform liquid-liquid extraction separating RNA from complex biological samples, and generates an aqueous phase containing pure, high-quality RNA [95].

During sample and TRIzol® homogenization, the cells are lysed. Despite cellular disruption and cell component solubilizing, TRIzol® maintains the RNA integrity due to high RNase activity inhibition. Adding chloroform to the lysate induces component phase separation. RNA is restrained in the aqueous phase, while the DNA and proteins are extracted into the organic phase. Isopropanol precipitates RNA from the aqueous phase, and the precipitate is then washed to remove impurities [95, 96]. The precipitated RNA is further purified using the RNeasy MinElute CleanUp Kit (Qiagen) prior to RNA concentration and purity measurements. Pure RNA samples displays a 260/280 ratio of ~2,0 and a 260/230 ratio in the range of 1,8–2,2. Lower ratios may indicate the presence of co-purified impurities [75].

The TRIzol® reagent contains possible hazardous compounds. Phenol is toxic and corrosive, while guanidine isothiocyanate is an irritant. Proper handling in a closed fume and with appropriate personal safety equipment is recommended [96].

The *WRAP53* knockdown efficiency in each cell line was verified at protein level by Western blot analyses and at mRNA level by real-time quantitative PCR (qPCR) (KAPA SYBR® FAST, KAPA Biosystems) [92]. In Western blot analysis, proteins are based on size and charge separated by gel electrophoresis, blotted onto a membrane and detected using labeled specific antibodies (WRAP53 C2 antibody, diluted 1:3000, Innovagen). Housekeeping proteins (here β-actin) and protein-specific antibodies (Monoclonal Anti-β-actin antibody produced in mouse, diluted 1:5000, Sigma Aldrich®) are commonly used as loading controls to ensure proper Western blot interpretations.

qPCR is a PCR-derived method combined with a reverse transcription reaction to both quantify and amplify the target sequence and detect gene expression changes. Continuous quantification is achieved by the use of a fluorescent reporter molecule where the signal intensity is proportional to the amount of amplified product. The KAPA SYBR® FAST method uses the non-specific fluorescent dye SYBR Green I. Free SYBR Green I in solution emits a restricted fluorescent signal, but bound to

dsDNA the signal intensity increases drastically. By monitoring the intensity of the fluorescent signals, the gene expression levels are quantified [19].

The RNA samples returned from CCK were stored at –80°C. Sample aliquots were used to prepare dilutions of 50 ng/µl RNA to use in the microarray experiment. DNase/RNase free water was used as dilution agent. The dilution concentrations and purity were measured using the NanDrop® ND-1000 Spectrophotometer the same way as described in Appendix A, but with the RNA application (RNA-40). The dilutions were stored at –80 °C.

3.4.4. Microarray-based gene expression profiling

The gene expression analyses was performed using the One-Color Microarray-Based Gene Expression Analysis and SurePrint G3 Hmn GE 8x60K Microarrays (Agilent Technologies, protocol in Appendix C). RNA from the siC and siWRAP53#2 transfected cell cultures were analyzed in biological triplicates, while RNA from the NTC cell cultures were analyzed in technical duplicates only.

Gene expression can be quantified measuring the mRNA transcript or protein levels by various techniques. DNA microarrays are tools allowing fluorescent labeled complementary RNA (cRNA) transcripts to hybridize to gene-specific complementary DNA (cDNA) or synthetic oligonucleotide probes, determining the relative expression levels of their corresponding genes. The intensity of the fluorescent signal is proportional to the relative abundance of specific cRNAs in the hybridized sample, which further reflects the occurrence of the corresponding mRNAs in the original sample. Large-scale and genome-wide transcript profiling was introduced in the late 1990s [19], and today several suppliers offer gene expression analysis tools where the expression levels of hundreds and thousands of genes are quantified simultaneously. The DNA microarrays make it possible to generate gene expression snapshots at different sampling times and conditions, answering the question of which genes are differently expressed between the analyzed samples.

The 8x60K microarrays used in the current study are designed by Agilent Technologies and printed using the Agilent SurePrint technology. Each array contains eight SurePrint sub-arrays consisting of 60 000 *in situ* synthesized oligonucleotide probes, each gene represented by several probes. The SurePrint technology generates 60 nucleotides long probes immobilized to fixed positions on surface-coated glass wafers. *In situ* probe synthesis prints the oligonucleotides directly onto the glass

microarray surface, and the coating strongly binds the probes to the array. The SurePrint technology prints thousands of probes at once, one nucleotide at a time, using the phosphoramidite chemistry [97]. Phosphoramidites are nucleotide analogs where the nucleotide reactive hydroxyl, phosphate and amine groups are chemically blocked by protective groups like dimethoxytrityl (DMT), methyl (Me) and isopropyl (iPr) to prevent undesired side reactions (figure 10), generating only the desired synthesized products. To link phosphoramidites monomers to oligonucleotide probes, the blocking groups are selectively removed. In the deprotection step, the DMT group is removed from the 5' hydroxyl group and the reactivity necessary to add the next nucleotide is reinduced. The iPr group protecting the 3' end of the adding nucleotide then departs and creates an intermediate that couples to the growing oligonucleotide. The last cycle step is oxidation to stabilize the phosphate linkage [97, 98], and the cycle is in total repeated 60 times to complete the oligonucleotides. The probes are then permanently bound to the microarray surface, and the surface is treated to reduce the amount of background signals [97].

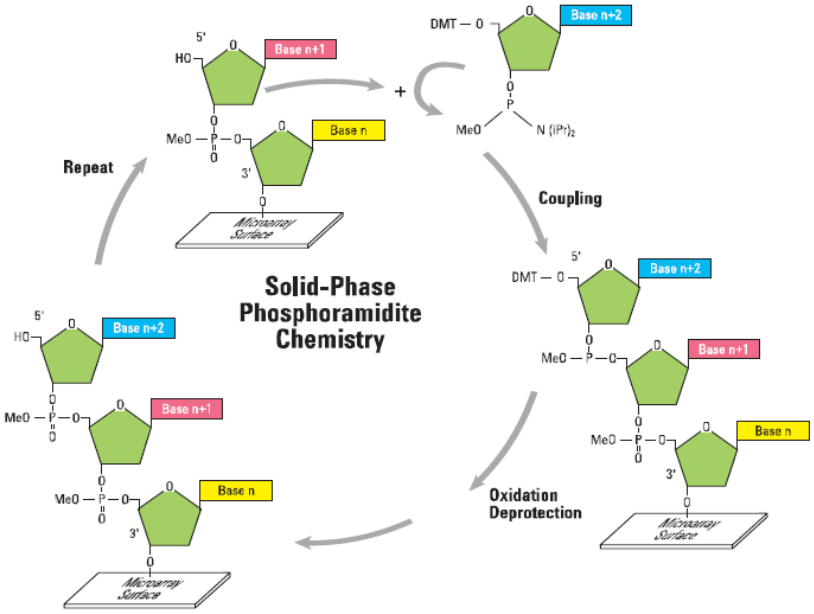


Figure 10: SurePrint technology *in situ* oligonucleotide probe synthesis by phosphoramidite chemistry. Removal of the DMT and iPr groups reinduces the reactivity necessary to couple phosphoramidite monomers into oligonucleotide probes. An oxidation step stabilizes the phosphate linkage connecting the monomers [97].

Microarray-based analyses are multi-step procedures where user-induced, sample and procedure variations might affect the microarray data. To eliminate such non-biological variation, positive controls are included in the analyses. The first step in the One-Color Microarray-Based Gene Expression Analysis (Agilent

Technologies) is to add One-Color Spike Mix (Agilent Technologies), a positive control consisting of ten different *in vitro* synthesized polyadenylated transcripts generated from the Adenovirus *E1A* gene, to the samples. The positive control transcripts hybridize to complementary probes on the array and reduces the non-biological variation in microarray data processing [99].

The fluorescent labeled cRNA hybridizing to the microarray is created in a two-step process of cDNA and cRNA synthesis. An oligo(dT)-T7 promoter primer is incorporated by hybridizing to the mRNA poly-A tails. The mRNA is then converted to cDNA in a reverse transcription reaction. After second strand cDNA synthesis, a T7 RNA polymerase creates labeled cRNA by incorporating Cyanine 3 (Cy3)-labeled cytosines (figure 11) [100]. The T7 RNA polymerase is isolated from the T7 bacteriophage and is commonly in used in *in vitro* RNA synthesis. The T7 RNA polymerase is extreme promoter-specific and catalyzes RNA synthesis by binding to the T7 RNA polymerase-specific binding site in the synthetic T7 promoter primer [73].

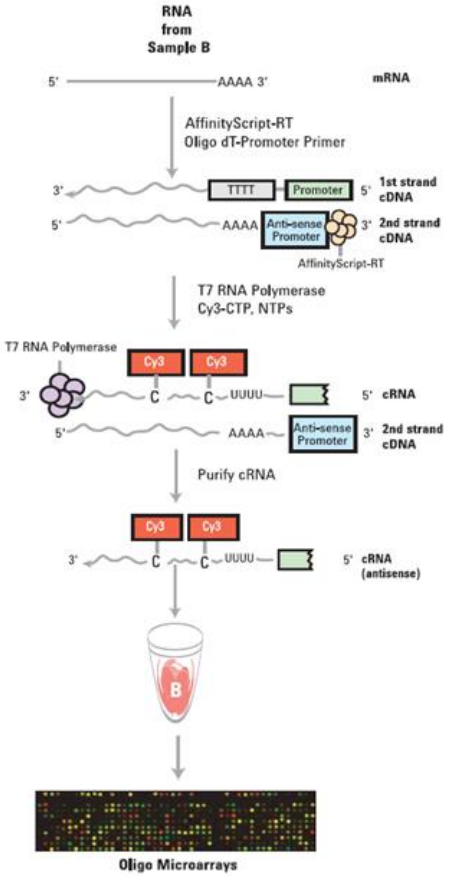


Figure 11: Schematic presentation of the cRNA synthesis in the One-Color Microarray-Based Gene Expression Analysis (Agilent Technologies) [100].

The amplified, labeled cRNA has to be purified prior to hybridization. The purification step was performed using the RNeasy® Mini Kit (Qiagen), a spin column method where the RNA-binding silica membrane has a binding capacity of 100 µg RNA. High-salt buffers and ethanol was added to optimize the binding conditions, and contaminants and excess fluids were washed away. The pure membrane-bound cRNA was then eluted in DNase/RNase free water [101]. The cRNA concentration and purity was measured using the microarray RNA application in the NanoDrop® ND-1000 Spectrophotometer. To proceed to the microarray hybridization step, a

cRNA yield of > 1,65 µg and specific activity of > 9,0 pmol Cy3 per µg cRNA are required [100]. The cRNA quantifications were also used to calculate the required sample volumes in the same step (Appendix C, step 4).

The purified cRNA was treated in a pre-hybridization reaction to ensure high-quality microarray data. The cRNA was mixed with a 10x Blocking Agent (Agilent Technologies) that minimizes non-specific array binding, and a 25x Fragmentation Buffer (Agilent Technologies) that fragments the cRNA to optimal sized targets to microarray hybridization [102]. After adding the 2x Hi-RPM Hybridization Buffer (Agilent Technologies), the samples were applied to the microarrays and hybridized for 17 hours. The cRNA fragments hybridize to their complementary probes and are in this way permanently attached to the array surface. All unbound compounds are removed in the wash step, creating microarrays of specific hybridized cRNA fragments ready to be scanned [100].

When scanning (Microarray Scanner with SureScan High Resolution Technology, Agilent Technologies) the microarrays, a laser beam irradiates the hybridized cRNA fragments and excites the Cy3 dye. The intensity of the emitted fluorescent light is detected and correlated to the expression levels of the cRNA corresponding genes. Cy3 conjugates are exited maximally at 550 nm while the emission peak maximum is detected at 570 nm [19]. The detected intensity values are then transformed to digital images of the microarrays.

After the microarrays are scanned, the Feature Extraction software (v10.7.3.1, Agilent Technologies) reads and processes the raw microarray images. The software finds and places the microarray grids, discards outlier pixels, determine probe intensities and ratios, flags outlier probes and performs statistical calculations generating quality control sample reports [103].

3.4.5. Gene expression data analysis

DNA microarray experiments generate enormous data amounts that have to be pre-processed prior to further analysis. The GeneSpring GX software (v12.0, Agilent Technologies) was used to normalize and filter the data. The normalized signal values were $\log_2$ transformed, baseline corrected and 75th percentile shift normalized, as recommended by the vendor. Normalization minimizes the impact of non-biological variation like unequal cRNA quantities, differences in inter-chip hybridization and differences between manufactured chips to expose the actual biological differences. In

75th percentile normalization, the 75th percentile signal intensity is calculated for each array and subtracted from all the expression values on the array. Baseline transformation does not affect the statistical analyses, but improves the data visualization in sample plot and map presentations in GeneSpring [104].

Sample quality control is a GeneSpring application where poor quality samples and probes can be eliminated. A principal component analysis (PCA) compares sample gene expression profiles, and shows variations and unknown trends in the dataset. The principal components are vectors capturing the data variance, and the samples are grouped according to gene expression similarities. Samples representing equal experimental conditions should be similar to each other and group closely in the PCA plot. Divergence from this alignment might be due to poor sample quality or actual biological variation. The PCA plot allows evaluation of deviating samples, the extent of variance accepted and if ambiguous samples should be eliminated prior to analysis [104, 105]. Probe filtering is as well a part of the data quality control and eliminates probes with unreliable expression measurements. This process excludes probes representing genes not expressed at significant levels in any of the samples, generating a list of quality probes for further statistical analysis [105].

Following normalization and quality control, the microarray data was analyzed using a combination of the Significance Analysis of Microarrays (SAM) and the Ingenuity Pathways Analysis (IPA) (Ingenuity® Systems). SAM[6] is an R-package, Excel add-in application which by gene-specific t-tests determines if gene expression alterations are statistically significant [106]. The microarray data was analyzed using the SAM two class unpaired test, grouping gene expression data from the NTC, siC and siWRAP53#2 transfected cells within each cell line at the different time points, investigating the genes with different expression patterns between the groups. The false discovery rate (FDR) estimates define the expected fraction of false positives among the significant results [107], i.e. the risk of falsely rejecting the null hypothesis (type I error) [86]. FDR estimates up to approximately 5% are generally accepted. Overall, differences in the gene expression pattern of the NTC cells and siC transfected cells within each cell line should be negligible. Comparison of the gene expression patterns in the siC and siWRAP53#2 transfected cells should on the other hand view the *WRAP53* knockdown effect.

[6] Available from http://www-stat.stanford.edu/~tibs/SAM/

The statistical significant differentially expressed genes obtained by SAM were further analyzed using the IPA software. IPA is a web-based[7] bioinformatics tool that allows complex data analyses and increased data understanding by the view of current molecular interactions, biological functions and diseases. The cellular importance of gene expression changes might as well be analyzed, as IPA predicts the downstream biological effects of such changes, an application used to analyze the microarray data in the current study. The IPA core analysis relates the experimental dataset molecules to the information in the Ingenuity Knowledge Base and identifies the signaling and metabolic pathways, networks of interconnected molecules, and biological functions and disease states most significant to the dataset of interest [108]. The results are provided with p-values, FDR values and ratios. The p-values are calculated by the Fisher`s exact test and estimate if the association between a group of molecules in the dataset and a given biological function in the Ingenuity Knowledge Base occur randomly by chance. Smaller p-values indicate more significant associations, and p-values $\leq 0{,}05$ were considered statistically significant. The FDR values are calculated by the Benjamini-Hochberg (B-H) method adjusting p-values by multiple testing [109, 110], and the FDR cut off was set to 5%. The ratios are the number of molecules from the dataset of interest represented in different pathways relative to the total number of molecules in the pathway [111].

IPA requires input gene lists that merely consist of unique, characterized genes. To identify the genes represented by the probes detected as statistically significant by SAM, the Stanford Microarray Database SOURCE[8] was used. SOURCE unifies probe data from different microarray platforms (Affymetrix, Agilent, Heebo/meebo, Illumina) and relates imported probe identification lists to the specific probe corresponding genes [112].

[7] Available from www.ingenuity.com
[8] Available from http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch

4. Results

4.1. *WRAP53* mutation analysis

To investigate the importance of somatic *WRAP53* mutations in breast cancer, the tumor samples in the ULL cohort were screened for *WRAP53* mutations by sequencing the entire *WRAP53* coding region (exon 2–11), start exon 1β and the exon flanking intronic regions (30 bps). *WRAP53* sequence alterations were detected in accordance with the NCBI *WRAP53* Reference Sequence NG_028245.1 [85] using the SeqScape software (v2.7, Applied Biosystems). A complete overview of the findings diverging from the *WRAP53* Reference Sequence is presented in Appendix D, table 10.

The *WRAP53* primers specificity was tested prior to the mutation analysis by PCR amplification of MDA-MB-231 genomic DNA and separation of the PCR products by agarose gel electrophoresis. The primer pairs generated specific PCR amplification products of expected size (table 1) and in sufficient quantities, and were further used in the mutation analysis.

To ensure high-quality sequencing results, the *WRAP53* fragments were sequenced in both forward and reverse direction based on independent PCR amplifications prior to the direction-specific cycle sequencing reactions. A complete *WRAP53* mutation screening of the ULL tumor samples therefore required nearly 4600 sequencing reactions and subsequent electropherograms to be interpreted. Approximately 70% of the sequenced fragments were covered by complete forward and reverse high-quality sequences, while the remaining part was covered by forward, reverse or a combination of the latter sequence directions, still generating result of adequate quality. The detected *WRAP53* sequence alterations in the ULL tumor samples were confirmed by the occurrence in both forward and reverse sequencing direction from the two independent PCR products. By analyzing a selection of tumor corresponding blood samples, known germ-line variation were validated, whereas new germ-line variants were distinguished from somatic mutations (*WRAP53* fragments analyzed in blood are in Appendix D, table 10 marked *).

4.4.1. Sequence alterations in the *WRAP53* gene

Compared to the NCBI *WRAP53* Reference Sequence NG_028245.1 [85], one or more *WRAP53* sequence alterations were detected in 86 out of the 175 (49%)

analyzed ULL tumor samples. The majority of the alterations were single base substitutions (in total 154 detected substitutions) of which many (109/154, 70%) were non-synonymous and therefore predicted to induce amino acid exchanges. The remaining (45/154, 30%) were synonymous, silent alterations where the amino acid order and protein structure remains unchanged (table 3) [19]. Indel mutations[9] were detected in only two of the cases (ULL-T-142 and ULL-T-253).

Five different base substitutions were detected in the coding region of *WRAP53*; (i) R68G, (ii) F150F, (iii) A436A, (iv) A522G and (v) A522A (table 3), of which the R68G and A552G are non-synonymous alterations resulting in amino acid substitutions from arginine and alanine to glycine, respectively. The same alteration found in tumor DNA was also detected when analyzing the corresponding blood samples, revealing that the base substitutions were germline and not somatic alterations.

Single nucleotide substitutions were also detected in the non-coding start exon 1β and the flanking intronic sequences (table 3). Of the sequence alterations in the non-coding *WRAP53* regions, the G>C base change in exon 1β occurred most frequently (4/154, 2,5%). The occurrence of the variant in exon 1β was verified in blood sample from the same patients, suggesting a common polymorphism. The question whether also the rare intronic variants were germline or somatic alterations were investigating using the SNP500Cancer [90] and NCBI dbSNP [91] databases, due to lack of patient blood samples. The databases classified all current registered *WRAP53* alterations as SNPs, i.e. normal population variations. Two possible novel intronic SNPs (intron 2 and 3) were identified, not previously registered in the current databases.

Intronic alterations might affect mRNA splicing if occurring in conserved splice site consensus sequences [19]. At the Department of Genetics, the criterion to define somatic sequence alterations as potential splice mutations requires an alteration localized within the two first bps prior to or following the coding regions. According to this criterion, none of the *WRAP53* intronic alterations found in the ULL tumor samples were classified as mRNA splice-affecting alterations.

Somatic mutations in the *WRAP53* gene do not seem to be a common event in breast tumorgenesis, although there is a little uncertainty concerning the indels, which

[9] Indel mutation is a generic term used to describe DNA insertions and deletions [19].

will be addressed beneath. Thus, the focus will further be on the SNPs in the *WRAP53* coding region, R68G, F150F and A522G, and the non-coding c.–245G>C (figure 12), all detected in at least two breast tumors in the ULL series of patients. The occurrence and disease related effect of the four SNPs was investigated by exploring the association to clinical, pathological and molecular parameters.

Table 3: Detected *WRAP53* single base substitutions in the ULL tumor samples. Alterations verified in blood samples are marked *. Chr = chromosome, ¤ = possible novel *WRAP53* SNPs not registered in the SNP500Cancer and NCBI dbSNP databases, - = data not available.

Localization	Chr position	Base change	Coding description	Codon change	Amino acid	Protein change	rs number
Exon 1β*	17, 7591722	G>C	c.–245G>C	-	-	-	rs17883670
Exon 2*	17, 7592168	C>G	c.202C>G	CGG>GGG	Arg>Gly	p.R68G	rs2287499
Exon 3*	17, 7592560	C>T	c.450C>T	TTC>TTT	Phe>Phe	p.F150F	rs2287498
Exon 10	17, 7606350	T>C	c.1308T>C	GCT>GCC	Ala>Ala	p.A436A	rs34016213
Exon 11*	17, 7606722	C>G	c.1565C>G	GCG>GGG	Ala>Gly	p.A522G	rs7640
Exon 11*	17, 7606723	G>A	c.1566G>A	GCG>GCA	Ala>Ala	p.A522A	rs148329158
Intron 2	17, 7592527	C>G	c.432–15C>G	-	-	-	¤
Intron 3	17, 7592657	G>A	c.530+17G>A	-	-	-	¤
Intron 5	17, 7604174	C>T	c.731+27C>T	-	-	-	rs138634236
Intron 6	17, 7604965	C>T	c.823–10C>T	-	-	-	rs117192546
Intron 8	17, 7606031	G>A	c.1165–30G>A	-	-	-	rs149142873

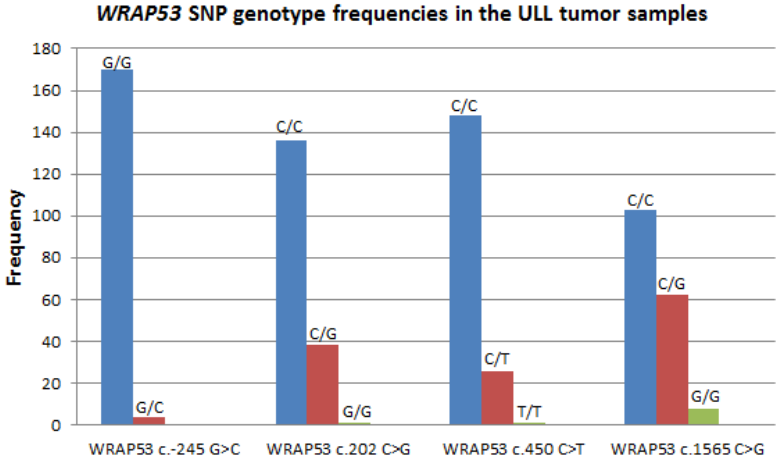


Figure 12: Frequency plot presenting the genotype distributions of the non-coding *WRAP53* c.–245G>C and the coding *WRAP53* c.202C>G (R68G), c.450C>T (F150F) and c.1565C>G (A522G) SNPs detected in the ULL tumor samples.

4.4.2. *WRAP53* indel sequence variations

In two of the ULL tumor samples, ULL-T-142 and ULL-T-253, *WRAP53* indel sequence alterations were detected. The indels were not registered in the

SNP500Cancer or NCBI dbSNP databases, indicating identification of possible novel sequence alterations. The alterations were localized close together in exon 11, suggesting a potential functional region of *WRAP53* that may be a target of somatic mutations. Sequencing the blood sample (ULL-B-253) revealed the same indel alteration as the one uncovered in the tumor sample (ULL-T-253) (figure 13), excluding the idea of this *WRAP53* indel alteration as a cancer-related somatic mutation. The alteration is instead a *WRAP53* germline variant present in all body cells of this patient. Blood samples from the ULL-T-142 patient do not exist, and it is not possible to determine whether the indel alteration is of somatic or germline origin.

The detected indel alterations were complex changes occurring in a region with a common SNP variant (A522G), which complicated the identification and annotation of the indels. Also, it was not possible to determine if the A522G SNP was present in these samples in addition to the indel alterations, and consequently the SNP in the two samples are not available and excluded from further statistical analyses.

Annotation of the ULL-T-142 sample predicts two possible frameshift alterations according to the presence or absence of the A522G SNP; (i) c.1566_1567insG and (ii) c.1564_1567delCGinsGGG, respectively. In the ULL-T-253 sample, the alteration was predicted a frameshift caused by c.1565_1568delGC. According to the NCBI *WRAP53* Reference Sequence NG_028245.1 [85], all these annotated indels introduce premature stop codons generating truncated gene products. Since the ULL-T-253 indel was identified as a germline alteration, there is a possibility that the ULL-T-142 indel might be a germline variant as well, but lack of patient blood samples makes this undeterminable. There were no obvious similarities in clinicopathological and molecular characteristics between the two patients.

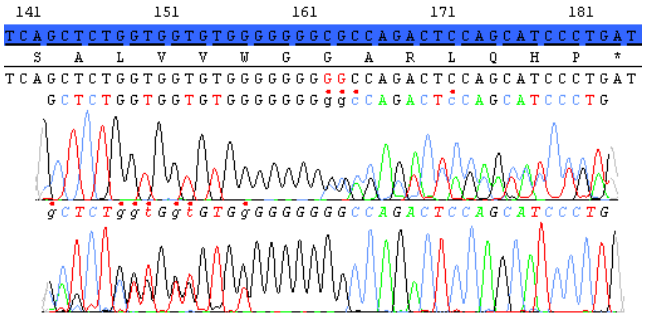


Figure 13: SeqScape electropherogram presenting the ULL-T-253 *WRAP53* indel alteration in exon 11. Deletions and insertions result in double peaks in both forward (normal) and reverse (italic) sequencing direction. The red dots are SeqScape flags due to interpretation uncertainties.

4.4.3. *WRAP53* SNPs in linkage disequilibrium

In 2007, Garcia-Closas et al. introduced the association between the *WRAP53* R68G SNP and the risk of developing ER negative breast cancer. The R68G SNP was observed to be in LD with the F150F SNP [65], an association of interest to validate in the ULL tumor samples.

The pair-wise association between the R68G, F150F, A522G and 1β c.–245G>C SNPs were investigated using the Chi-square and Fisher`s exact tests. Due to few observations in the groups of R68G and F150F minor homozygous genotypes (G/G and T/T, respectively), these observations were merged with the heterozygous genotypes within each SNP prior to the statistical analyses. In the pair-wise SNPs association analyses, all three A522G nucleotide variants were included, but were then merged like the R68G and F150F nucleotide variants for the subsequent analyses. The null hypothesis indicated no association between the occurrences of the different *WRAP53* SNPs, while the alternative hypothesis indicated a significant association. Statistical significant associations were observed between the R68G and F150F, R68G and A522G (figure 14), and F150F and A522G SNPs, all with p-values < 0,001. This implies a non-random occurrence of *WRAP53* R68G, F150F and A522G in the breast cancer patients, where the homozygous minor genotypes and the heterozygous genotypes respectively are inherited together.

Crosstabulation WRAP53 R68G and A522G

			WRAP53 A522G			Total
			C/C	C/G	G/G	
WRAP53 R68G	C/C	Count	102	30	3	135
		% within WRAP53 c.202 C>G	75,6%	22,2%	2,2%	100,0%
	C/G + G/G	Count	1	32	5	38
		% within WRAP53 c.202 C>G	2,6%	84,2%	13,2%	100,0%
Total		Count	103	62	8	173
		% within WRAP53 c.202 C>G	59,5%	35,8%	4,6%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	65,949[a]	2	,000
Likelihood Ratio	74,428	2	,000
Linear-by-Linear Association	60,966	1	,000
N of Valid Cases	173		

a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is 1,76.

Figure 14: The Chi-square test result indicates a statistical significant association between the *WRAP53* R68G and A522G SNPs (p < 0,001).

Genetic linked loci defines haplotypes, i.e. alleles on the same chromosomal segment inherited together as blocks [19]. The Haploview software (v4.2, Broad Institute) was used to further investigate whether the associated *WRAP53* SNPs were genetic linked. The degree of genetic linkage within the dataset is visualized in a LD plot (figure 15 A) and specified by LOD scores (table 4). The Haploview result strongly indicated that the *WRAP53* R68G, F150F and A522G SNPs occur in LD as a haplotype block, supporting the findings in the Chi-square and Fisher`s exact test analyses of the same SNPs. Haploview further confirmed that the HW equilibrium was fulfilled, meaning there is no selection of specific *WRAP53* genotypes in the breast cancer patients. The *WRAP53* haplotype block generated by Haploview shows the marker SNPs, the haplotypes and their frequencies, and reveals that CCC is the most frequent genotype among the ULL breast cancer patients (figure 15 B).

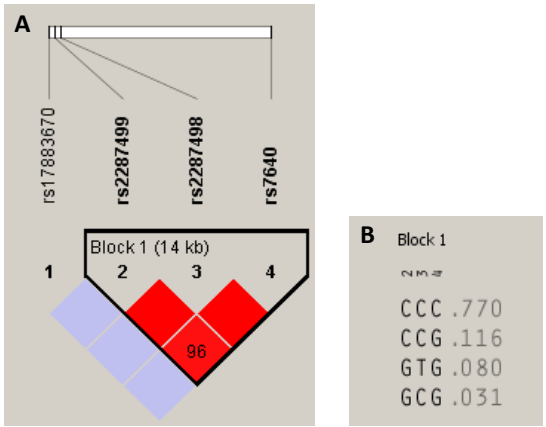


Figure 15: Color scheme based on D' and LOD score values. Bright red D' = 1 and LOD > 2, light blue D' = 1 and LOD < 2. Number in the squares are D' values (values of 1,0 are not shown). The physical position of each SNP is shown above the plot (A). The *WRAP53* haplotype block with marker numbers, the haplotypes and the haplotype frequencies (B).

Table 4: LOD scores from the Haploview (v4.2, Broad Institute) haplotype analysis

***WRAP53* SNPs (rs numbers)**	**LOD score**
rs2287498 (F150F) and rs7640 (A522G)	12,6
rs2287499 (R68G) and rs7640 (A522G)	17,0
rs2287499 (R68G) and rs2287498 (F150F)	23,9
rs17883670 (c.–245G>C) and rs2287499 (R68G)	0,27
rs17883670 (c.–245G>C) and rs2287498 (F150F)	0,18
rs17883670 (c.–245G>C) and rs7640 (A522G)	0,56

4.4.4. Association of *WRAP53* SNPs to clinicopathological and molecular data

The *WRAP53* SNPs R68G, F150F, A522G and 1β c.–245G>C were individually explored in order to reveal their associations to clinical, pathological and molecular patient parameters. The statistical analyses was performed using the Chi-square and Fisher`s exact tests. The SNPs occurrence were tested against the following current clinicopathological parameters in breast cancer; (i) age of disease onset (cases divided in two groups, < > 55 years, due to pre and post menopause since endogenous hormonal levels are a breast cancer risk factor), (ii) lymph node status (positive or negative according to the presence of lymph node metastasis), (iii) tumor size (T1–T4[10]) [113], (iv) histological grade (G1–G4[11]) [113], (v) *TP53*-status (wt or mutated), (vi) ER status (positive or negative) and (vii) breast cancer subtypes (luminal A, luminal B, ERBB2 positive, basal-like and normal breast-like). In addition, the molecular parameter of cellular WRAP53 protein localization (nuclear, cytoplasmic or combined) was included to investigate the functional SNPs effect in accordance with unpublished findings associating cellular WRAP53 localization with breast cancer prognosis (Langerod et al., unpublished data). Complete clinical information was not available to all samples. The null hypothesis indicated no association between the current *WRAP53* SNPs and the clinical, pathological and/or molecular parameters, while the alternative hypothesis indicated a significant association.

No statistical significant results were obtained investigating the association of the *WRAP53* SNPs (R68G, F150F, A522G and 1β c.–245G>C) with the patient parameters listed above, except two findings concerning cellular WRAP53 protein localization. The heterozygous *WRAP53* c.–245G>C genotype (G/C) was found associated with nuclear WRAP53 localization (p=0,052) (figure 16). This finding is intriguing since nuclear WRAP53 localization is presumed associated with favorable prognosis in breast cancer patients compared to cytoplasmic localization (Langerod et al., unpublished data). The R68G SNP was associated to cellular WRAP53 localization (p=0,036) combining the nuclear and cytoplasmic protein distributions. Positive nuclear and negative cytoplasmic WRAP53 localization improves breast cancer prognosis in contrast to the counterpart of negative nuclear and positive

[10] T1=primary tumor < 2,0 cm in diameter, T2=primary tumor > 2,0 < 5,0 cm in diameter, T3=primary tumor > 5,0 cm in diameter and T4=primary tumor independent of size, but the tumor infiltrates skin or breast tissue [113].

[11] G1=well differentiated (low grade), G2=moderately differentiated (intermediate grade), G3=poorly differentiated (high grade) and G4=undifferentiated (high grade) [113].

cytoplasmic localization. The combinations in between by positive or negative nuclear and cytoplasmic protein localizations displays similar, intermediate prognostic impact (Langerod et al., unpublished data).The R68G major homozygous (C/C) genotype is fairly even distributed between the four annotated groups of WRAP53 localization. The heterozygous (C/G) and minor homozygous (G/G) genotypes are on the other hand primarily distributed between negative or positive nuclear and cytoplasmic WRAP53 localization (p=0,036) (figure 17).

Crosstabulation WRAP53 c.-245G>C and nuclear WRAP53 protein localization (IHC)

			Nuclear WRAP53 protein localization		
			neg (0-5%)	pos (>5%)	Total
WRAP53 c.-245G>C	G/G	Count	67	59	126
		% within WRAP53 c.-245 G>C	53,2%	46,8%	100,0%
	G/C	Count	0	4	4
		% within WRAP53 c.-245 G>C	,0%	100,0%	100,0%
Total		Count	67	63	130
		% within WRAP53 c.-245 G>C	51,5%	48,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,389[a]	1	,036		
Continuity Correction[b]	2,518	1	,113		
Likelihood Ratio	5,930	1	,015		
Fisher's Exact Test				,052	,052
Linear-by-Linear Association	4,355	1	,037		
N of Valid Cases	130				

a. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,94.
b. Computed only for a 2x2 table

Figure 16: The Fisher`s exact test indicates a statistical significant association between the *WRAP53* c.-245G>C SNP and nuclear WRAP53 localization (p=0,052). Due to background staining, nuclear stain > 5% was considered positive results. IHC=immunohistochemistry.

Crosstabulation WRAP53 R68G and IHC WRAP53 protein localization (combined nucleus 5% and cytoplasm 5%)

			WRAP53 protein localization (nuclues 5% and cytoplasm 5%)				
			neg nucl / neg cyt	neg nucl / pos cyt	pos nucl / neg cyt	pos nucl / pos cyt	Total
WRAP53 R68G	C/C	Count	21	32	15	33	101
		Expected Count	24,9	27,2	13,2	35,7	101,0
		% within WRAP53 c.202 C>G	20,8%	31,7%	14,9%	32,7%	100,0%
	C/G + G/G	Count	11	3	2	13	29
		Expected Count	7,1	7,8	3,8	10,3	29,0
		% within WRAP53 c.202 C>G	37,9%	10,3%	6,9%	44,8%	100,0%
Total		Count	32	35	17	46	130
		Expected Count	32,0	35,0	17,0	46,0	130,0
		% within WRAP53 c.202 C>G	24,6%	26,9%	13,1%	35,4%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8,530[a]	3	,036
Likelihood Ratio	9,250	3	,026
Linear-by-Linear Association	,001	1	,975
N of Valid Cases	130		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 3,79.

Figure 17: The Chi-square test indicates a statistical significant association between the *WRAP53* R68G SNP and subcellular WRAP53 protein localization (p=0,036). Due to background staining, nuclear and cytoplasmic stain > 5% was considered positive results. IHC=immunohistochemistry.

4.4.5. *WRAP53* SNPs and survival analyses

The prognostic value of the four SNPs, *WRAP53* R68G, F150F, A522G and c.–245G>C were investigated using the Kaplan-Meier survival plot and the Logrank test. The null hypothesis indicated no association between the current *WRAP53* SNPs and breast cancer-specific survival, while the alternative hypothesis indicated a significant association.

The F150F SNP was the only *WRAP53* sequence alteration associated with breast cancer-specific survival (p=0,048) (figure 18), where the heterozygous (C/T) genotype seems to reduce the mean survival time with approximately 30 months (table 5). The same trend was observed for both the R68G (p=0,114) and A522G (p=0,133) SNPs, without showing statistical significance levels, but still indicating that the heterozygous and minor homozygous genotypes might be associated with reduced survival.

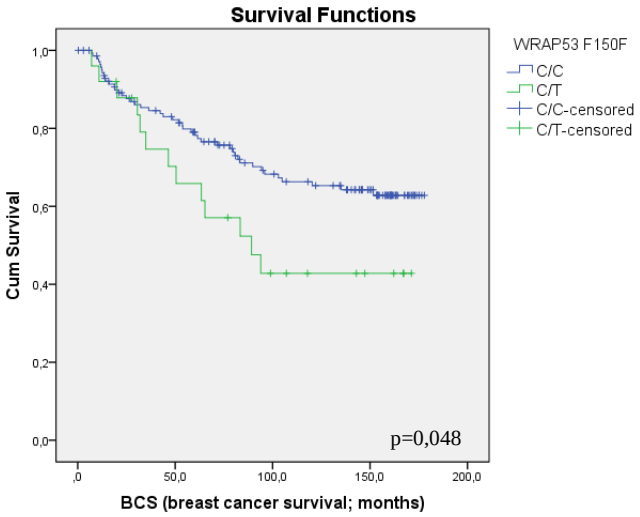


Figure 18: The Kaplan-Meier survival plot and Logrank test indicated a statistical significant association between the *WRAP53* F150F (c.450C>T) SNP and breast cancer-specific survival (p=0,048).

TP53 mutation status and ER status are significant and well established prognostic markers in breast cancer. The *WRAP53* SNPs significantly (F150F) or trend (R68G and A522G) associated with survival were further analyzed to investigate whether the SNP genotypes influenced breast cancer-specific survival differentially according to *TP53* mutation and ER status.

Both *TP53* mutation and ER status stratification revealed significant A522G genotype-dependent differences in survival. The heterozygous (C/G) and minor homozygous (G/G) genotype carriers were significantly associated with poorer

prognosis compared to the major homozygous (C/C) genotype carriers in wt *TP53* tumors (p=0,055), whereas the survival in *TP53* mutated tumors were not influenced (p=0,926) (figure 19 A and B). Interestingly, a similar result was seen by stratifying for ER status where reduced survival was associated with ER positive tumors in patients carrying the *WRAP53* heterozygous (C/G) and minor homozygous (G/G) genotypes (p=0,017), while survival in the ER negative tumors were not influenced (p=0,476) (figure 19 C and D). The R68G SNP survival analysis showed similar results by significantly reduced survival of heterozygous (C/G) and minor homozygous (G/G) genotypes compared to the major homozygous (C/C) genotype in wt *TP53* and ER positive tumors (p=0,030 and 0,004, respectively), while no significant difference in survival was observed in the *TP53* mutated and ER negative tumors (p=0,625 and 0,702, respectively).

The results from the F150F SNP survival analysis diverged compared to the A522G and R68G results. There was no statistical significant survival difference in wt *TP53* and *TP53* mutated tumors according to genotype (p=0,072 and 0,922, respectively). Despite the absent statistical significance, a possible trend was present in the wt *TP53* tumor patients indicating that heterozygous (C/T) and minor homozygous (T/T) genotype carriers seemed to be associated with poorer survival compared to the major homozygous (C/C) genotype carriers. By ER status stratification, the result was in accordance with the A522G and R68G outcomes. ER positive patients carrying the heterozygous (C/T) and minor homozygous (T/T) genotypes were significantly associated with poorer survival compared to the major homozygous (C/C) genotype carriers (p=0,040), while no significant survival difference was observed in the ER negative tumor patients (p=0,216) (table 5).

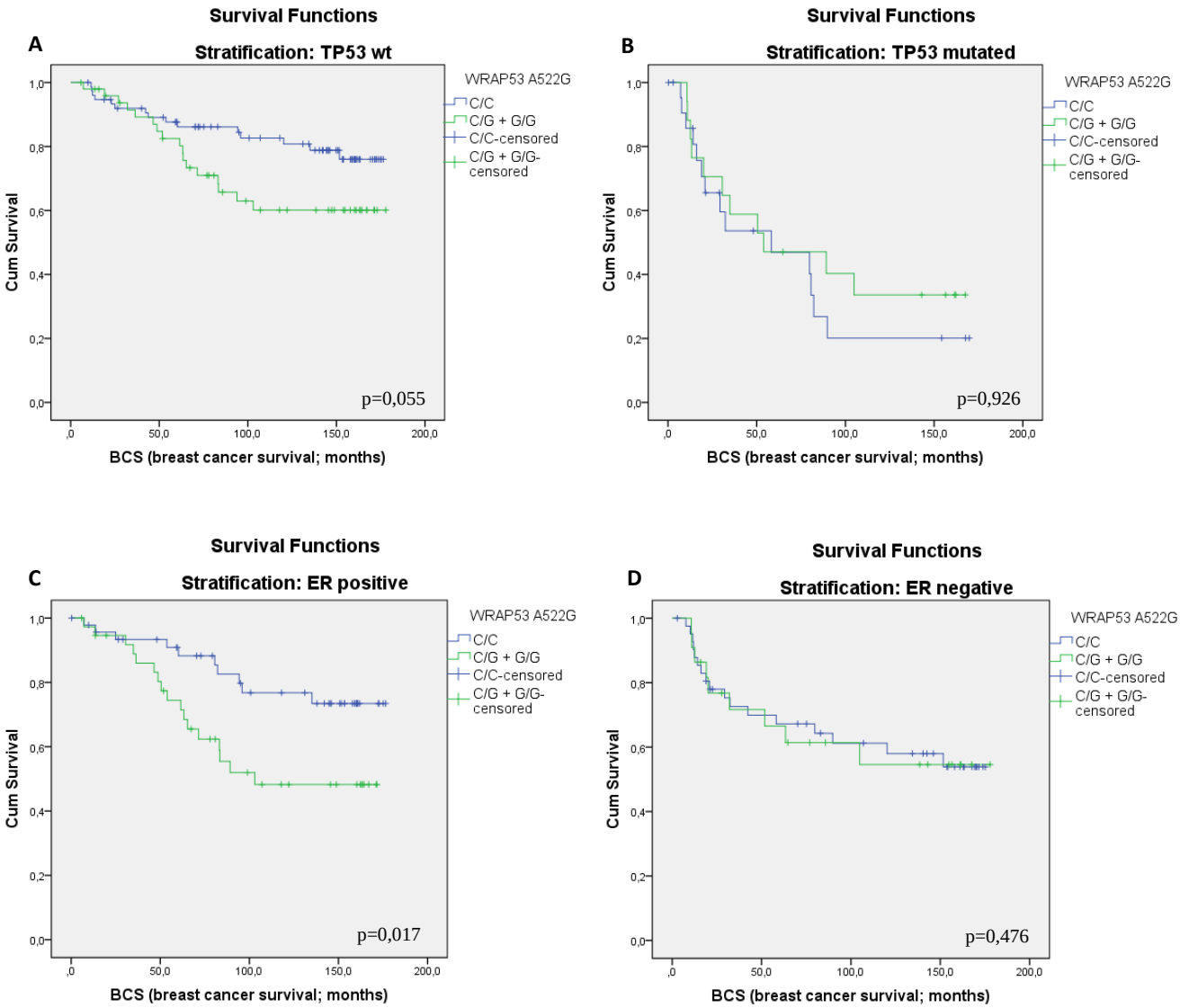


Figure 19: The Kaplan-Meier survival plots and Logrank tests indicated a statistical significant association between the A522G SNP and breast cancer-specific survival in wt *TP53* (A) and ER positive (C) tumors (p=0,055 and 0,017, respectively). No significant difference in survival was observed in *TP53* mutated (B) and ER negative (D) tumors (p=0,926 and 0,476, respectively).

Table 5: The effect of *WRAP53* SNPs on unstratified mean survival times and mean survival times stratified according to *TP53* mutation and ER status in the ULL tumor samples

***WRAP53* SNPs**	**Genotype**	**Mean survival time prior to stratification (months)**	**Mean survival time after stratification (months)**			
			ER status		***TP53* mutation status**	
			ER positive	**ER negative**	***TP53* wt**	***TP53* mutated**
R68G	C/C	131,9	144,7	113,9	148,6	73,5
	C/G+G/G	112,9	101,5	121,7	122,4	80,3
F150F	C/C	132,6	140,0	120,1	147,3	77,2
	C/T+T/T	101,7	102,9	72,5	117,1	70,0
A522G	C/C	134,5	148,1	117,0	150,8	68,7
	C/G+G/G	118,5	112,5	115,2	130,2	83,3

4.2. Gene expression study

To increase the knowledge about cellular *WRAP53* regulatory functions, gene expression analyses was performed in *WRAP53* normal expressing and *WRAP53* depleted breast cancer cell lines. The MCF-7 and MDA-MD-231 cell lines were cultured and transiently transfected with an siRNA targeting *WRAP53* exon 2. RNA was harvested from NTC cells, siC and siWRAP53#2 transfected cells in each cell line 40 and 72 hours after transfection. Gene expression differences according to the *WRAP53* knockdown within the respective cell lines were investigated using the SurePrint G3 Hmn GE 8x60K Microarray (Agilent Technologies).

4.2.1. *WRAP53* knockdown efficiency and RNA isolation

The MCF-7 and MDA-MB-231 *WRAP53* knockdown efficiency was measured quantifying the cellular WRAP53 protein and mRNA levels by Western blot and qPCR analyses, respectively. Normally, *WRAP53* is more abundantly expressed in the MDA-MB-231 than the MCF-7 cell line, visualized by stronger siC protein bands that reflects normal WRAP53 protein levels (figure 20). The *WRAP53* knockdown efficiency was most advantageous in the MCF-7 cell line, measured to 75% 72 hours after transfection. At the same time point in the MDA-MB-231 cell line, the knockdown efficiency was measured to 70% (figure 21). β-actin was stably expressed in both cell lines independently of siRNA transfections and time points (figure 20), indicating that the reduced WRAP53 protein levels were caused by the targeted *WRAP53* knockdown.

The purity of the TRIzol® isolated RNA samples was measured by 260/280 ratios within the range of 1,75–1,95 and 260/230 ratios within the range of 1,62–2,41.

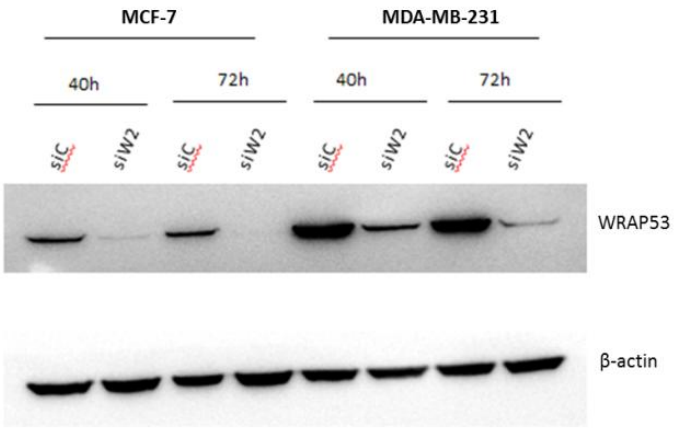


Figure 20: Western blot analysis quantifyed the WRAP53 protein levels in the siC and siWRAP53#2 transfected MCF-7 and MDA-MB-231 cell lines 40 and 72 hours after transfection. β-actin was used as loading control. *WRAP53* knockdown efficiency was determined using the Bradford assay (Bio-Rad Laboratories) and detected using the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific). Farnebo and colleagues, CCK (January 2012).

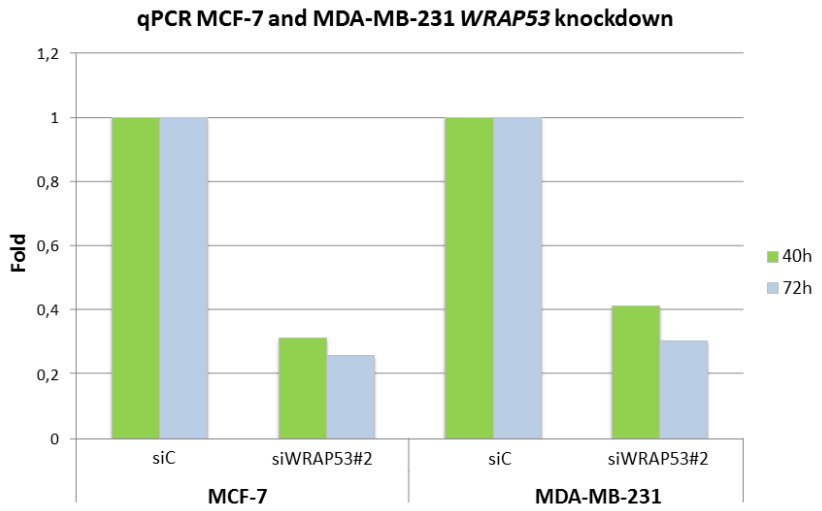


Figure 21: MCF-7 and MDA-MB-231 *WRAP53* knockdown efficiency measured 40 (green) and 72 (blue) hours after siRNA transfection using the KAPA SYBR® FAST qPCR (KAPA Biosystems). Farnebo and colleagues, CCK (January 2012).

4.2.2. Quantification and purity assessments of purified cRNA

High-quality cRNA samples are crucial to the success of an Agilent Gene Expression experiment. All samples had a cRNA yield within the range of 4,91–8,77 µg and a specific activity within the range of 9,38–14,44 pmol Cy3 per µg cRNA, all in accordance with the Agilent Gene Expression experiment procedure requirements (chapter 3.4.4) [100].

4.2.3. GeneSpring GX 12.0; preprocessing microarray data

The Feature Extraction software (Agilent Technologies) processed raw microarray images files were further processed using the GeneSpring GX 12.0 software. The microarray data was normalized using the 75th percentile shift method, and probe filtration generated a list of 35636 quality probes further included in the downstream gene expression data analyses.

The GeneSpring sample quality control revealed one sample (MDA-MB-231 siWRAP53#2 transfected cells, 40 hours, triplicate 1) (table 2) diverging from the sample majority localization in the PCA plot. The remaining corresponding replicates (MDA-MB-231 siWRAP53#2 transfected cells, 40 hours, triplicate 2 and 3) clustered together, indicating that biological variation probably was not the cause of divergence. To avoid unreliable measurements affecting the gene expression data, MDA-MB-231 siWRAP53#2 triplicate 1 was eliminated from the dataset.

Removing the divergent sample, the GeneSpring PCA plot clustered the MCF-7 and MDA-MB-231 cell lines separately in two distinct groups (figure 22). Variation was still present in the cell line-specific samples, but mainly distributed on the minor variance y- and z-axes components. Samples of similar experimental conditions are expected clustering together, but there was no clear trends of treatment-specific clustering within the cells lines.

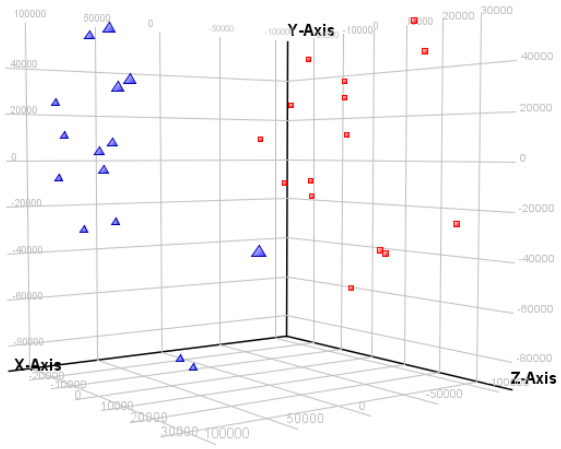


Figure 22: PCA plot visualizing microarray sample variations. Red squares equal MCF-7, while blue triangles equal MDA-MB-231. Cell line specificity dominates the clustering. The x-axis (component 1) captures 37,7%, the y-axis (component 2) 15,9% and the z-axis (component 3) 6,9% of the dataset variations.

4.2.4. Significance Analysis of Microarrays

To investigate the influence of *WRAP53* depletion according to MCF-7 and MDA-MB-231 gene expression alterations, the GeneSpring GX 12.0 preprocessed expression data was analyzed using SAM. SAM identifies and lists significant differentially expressed genes between groups of samples.

The SAM two class unpaired test analyzed differences in gene expression patterns between the (i) NTC cells and siC transfected cells, and (ii) the siC and siWRAP53#2 transfected cells within each cell line at both time points. There were no significant gene expression differences between the NTC cells and siC transfected cells at any time point, indicating that the siRNA transfection procedure itself did not influence with the gene expression patterns. Prospective gene expression differences in the siWRAP53#2 transfected cells are thus *WRAP53* knockdown-specific cellular responses.

Comparing gene expression patterns in the siC and siWRAP53#2 transfected cells within each cell line according to time points revealed considerable differences. In the MCF-7 cell line, the siWRAP53#2 transfected cells displayed 32 probes

(FDR=5,3%) representing genes of significant different expression levels compared to the siC transfected cells after 40 hours, all upregulated. 72 hours after transfection, the number of probes representing significant differentially expressed genes was expanded to 420 (FDR=0,5%), distributed in 406 upregulations and 14 downregulations (figure 23). The cellular *WRAP53* knockdown response in MDA-MB-231 seemed to occur more slowly than in MCF-7, showing no significant differences in gene expression patterns after 40 hours. By 72 hours, the *WRAP53* depletion effect in MDA-MB-231 was more prominent by significant altered expression levels of 453 (FDR=10,5%) gene-representing probes. *WRAP53* was among the significantly downregulated genes in the siWRAP53#2 transfected cells in both cell lines, and the only downregulated gene in MDA-MB-231, indicating efficient *WRAP53* knockdown.

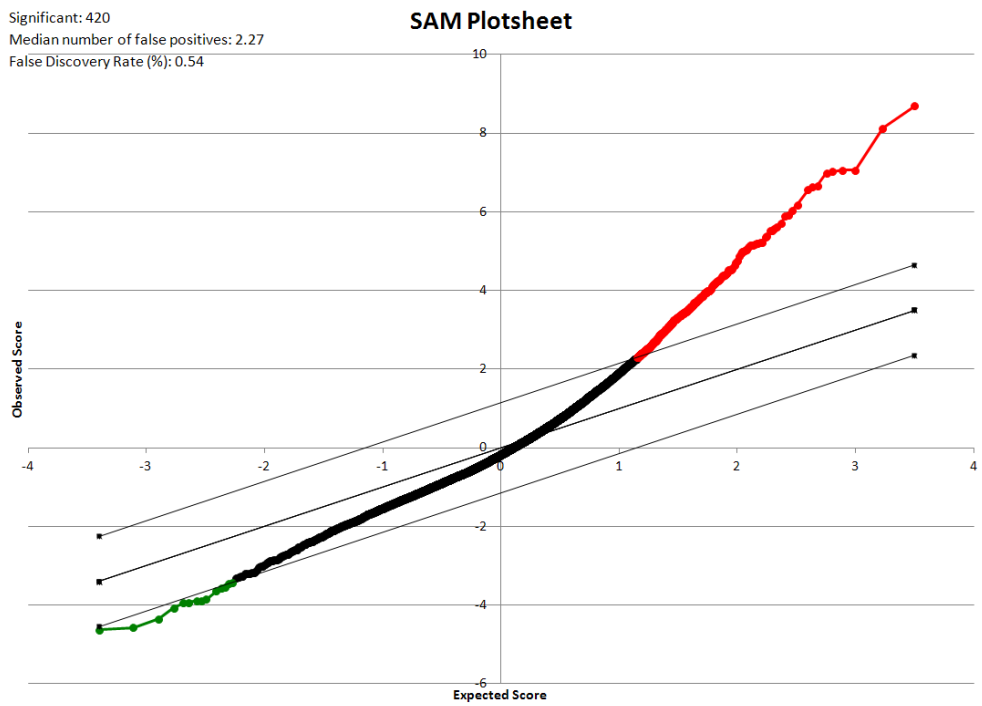


Figure 23: SAM plot visualizing the gene expression patterns in the MCF-7 siC and siWRAP53#2 transfected cells 72 hours after siRNA transfection. The red dots represent significant upregulated genes in the siWRAP53#2 transfected cells compared to the siC transfected cells. The green dots represent significantly downregulated genes in the siWRAP53#2 transfected cells compared to the siC transfected cells. The black dots represents genes expressed at equal levels in both groups. SAM estimated 420 probes representing genes with significantly altered expression levels (FDR=0,5%).

4.2.5. Pathway analyses

To investigate the impact of MCF-7 and MDA-MB-231 *WRAP53* depletion in accordance to biological functions, pathway analysis using IPA was performed. The SAM output lists of significant differentially expressed genes within each cell line were used in the pathway analyses. Prior to the analyses, the gene lists were processed using the Stanford Microarray Database SOURCE to identify probes of missing gene annotations to create lists exclusively containing distinctive, characterized genes. The pathway analyses p-value was calculated using the right-tailed Fisher`s exact test, and the B-H adjusted p-value for multiple testing was applied for gene lists with more than 100 genes. Due to overall restricted alterations in gene expression downregulations and limited cellular *WRAP53* knockdown response 40 hours after transfection, only significant upregulated genes at 72 hours were analyzed (gene lists in Appendix E). The top five most significant pathway analyses results of biological functions including diseases and disorders, molecular and cellular functions, canonical pathways and transcription factors are summarized in table 6 and 7.

The functional analyses identify the biological functions in the Ingenuity Knowledge Base that are most significant to the molecules in the dataset. The canonical pathway analysis identifies the pathways most significant to the dataset, and the association significance is measured in two ways: (i) a ratio of the number of molecules that map to the canonical pathway is displayed, and (ii) a calculated p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance. The transcription factor analysis in IPA predicts the transcription factors expected to be involved in regulation of the genes in the gene set. The Fisher`s exact test and B-H p-value threshold was set to 0,05, accepting a FDR of 5%, to identify biological functions significantly associated to the gene expression alterations in the gene dataset [114].

Table 6: Summary of pathway analysis (IPA) of MCF-7 significant upregulated genes (299 genes) 72 hours after transfection. B-H = Benjamini-Hochberg.

Diseases and disorders	**B-H p-value**	**#Molecules**
Cancer	$5{,}39\times10^{-9} – 6{,}38\times10^{-2}$	131
Gastrointestinal Disease	$3{,}09\times10^{-8} – 6{,}38\times10^{-2}$	75
Dermatological Diseases and Conditions	$1{,}59\times10^{-5} – 5{,}63\times10^{-2}$	56
Developmental Disorder	$1{,}59\times10^{-5} – 4{,}73\times10^{-2}$	34
Genetic Disorder	$1{,}59\times10^{-5} – 5{,}15\times10^{-2}$	88
Molecular and cellular functions	**B-H p-value**	**#Molecules**
Cell Death	$4{,}41\times10^{-7} – 6{,}38\times10^{-2}$	108
Cellular Movement	$8{,}25\times10^{-6} – 6{,}38\times10^{-2}$	75

Cellular Growth and Proliferation	$2,53\times10^{-5} – 6,25\times10^{-2}$	104
Free Radical Scavenging	$4,24\times10^{-5} – 6,04\times10^{-2}$	25
Cell Morphology	$2,10\times10^{-4} – 5,53\times10^{-2}$	65
Canonical pathways	**B-H p-value**	**Ratio**
VDR/RXR Activation	$7,92\times10^{-2}$	7/79 (0,089)
Aryl Hydrocarbon Receptor Signaling	$7,92\times10^{-2}$	9/145 (0,062)
p53 Signaling	$9,71\times10^{-2}$	7/95 (0,074)
Semaphorin Signaling in Neurons	$9,71\times10^{-2}$	5/52 (0,096)
Extrinsic Prothrombin Activation Pathway	$9,71\times10^{-2}$	3/16 (0,188)
Transcription factors	**p-value**	**#Target molecules**
ATF2	$6,85\times10^{-9}$	12
FOSL2	$1,26\times10^{-8}$	9
ATF3	$3,61\times10^{-8}$	10
TP53	$7,42\times10^{-8}$	45
TP63	$7,66\times10^{-8}$	16

Table 7: Summary of pathway analysis (IPA) of MDA-MB-231 significant upregulated genes (249 genes) 72 hours after transfection. B-H = Benjamini-Hochberg.

Diseases and disorders	**B-H p-value**	**#Molecules**
Dermatological Diseases and Conditions	$5,33\times10^{-5} – 7,48\times10^{-2}$	34
Genetic Disorder	$5,33\times10^{-5} – 7,48\times10^{-2}$	56
Cancer	$5,33\times10^{-5} – 7,48\times10^{-2}$	84
Gastrointestinal Disease	$5,33\times10^{-5} – 7,48\times10^{-2}$	59
Inflammatory Response	$1,19\times10^{-3} – 7,48\times10^{-2}$	45
Molecular and cellular functions	**B-H p-value**	**#Molecules**
Cellular Movement	$1,01\times10^{-5} – 7,48\times10^{-2}$	60
Cellular Growth and Proliferation	$1,54\times10^{-4} – 7,48\times10^{-2}$	74
Antigen Presentation	$2,04\times10^{-3} – 7,48\times10^{-2}$	23
Cell-To-Cell Signaling and Interaction	$6,32\times10^{-3} – 7,48\times10^{-2}$	40
Cellular Development	$8,67\times10^{-3} – 7,48\times10^{-2}$	65
Canonical pathways	**B-H p-value**	**Ratio**
Differential Regulation of Cytokine Production in Intestinal Epithelial Cells by IL-17A and IL-17F	$2,07\times10^{-3}$	5/23 (0,217)
Role of IL-17A in Psoriasis	$2,07\times10^{-3}$	4/13 (0,038)
Airway Pathology in Chronic Obstructive Pulmonary Disease	$8,71\times10^{-3}$	3/8 (0,375)
Hepatic Fibrosis / Hepatic Stellate Cell Activation	$2,39\times10^{-2}$	8/142 (0,056)
Role of IL-17A in Arthritis	$3,08\times10^{-2}$	5/60 (0,083)
Transcription factors	**p-value**	**#Target molecules**
FOXL2	$2,12\times10^{-10}$	11
RELA	$7,35\times10^{-7}$	17
ESR1	$2,79\times10^{-6}$	16
NFkB (complex)	$3,59\times10^{-6}$	22
TFAP2A	$3,28\times10^{-5}$	8

4.2.6. Significant mutually expressed MCF-7 and MDA-MB-231 genes

In order to understand the biological effects of *WRAP53* knockdown, the mutually expressed MCF-7 and MDA-MB-231 genes were investigated. The MCF-7 and MDA-MB-231 response to *WRAP53* depletion primarily occurred as gene expression

upregulations. The lists of distinctive, characterized MCF-7 and MDA-MB-231 upregulated genes (299 and 249 genes, respectively) 72 hours after transfection were used to identify mutually upregulated genes using a Venn diagram tool [115]. Venn diagrams visualizes the relations between different data sets [116], and 29 mutually upregulated genes were identified in the two cell lines (figure 24 and table 8). The mutually expressed genes were further included in the pathway analysis, and the results are summarized in table 9.

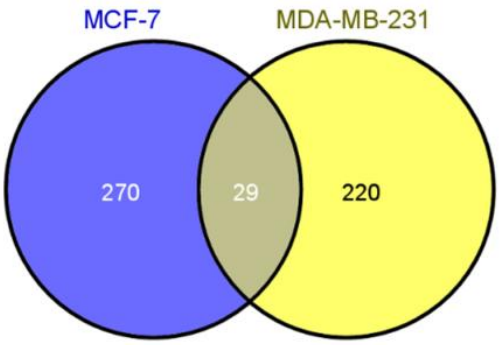


Figure 24: Venn diagram [115] visualizing the relation of mutually expressed upregulated genes (29 genes) in the MCF-7 and MDA-MB-231 cell lines 72 hours after transfection.

Table 8: Mutually significant MCF-7 and MDA-MB-231 upregulated genes 72 hours after transfection

Genes			
CACNG6	*EVPLL*	*MEF2C*	*THBD*
CCDC80	*FSIP2*	*MMP1*	*TM4SF1*
CLDN1	*GABARAPL1*	*NUPR1*	*TNFRSF11B*
CMAHP	*ITGB8*	*OSBPL5*	*TNFSF9*
COL20A1	*KLK6*	*SERPINB5*	*UCHL1*
CYB5RL	*LHPP*	*SLC13A3*	
CYP1B1	*LTB*	*SLC16A14*	
EDA2R	*MAP2*	*SOX4*	

Table 9: Summary of pathway analysis (IPA) of MCF-7 and MDA-MB-231 mutually significant upregulated genes (29 genes) 72 hours after transfection

Diseases and disorders	**p-value**	**#Molecules**
Cancer	$1{,}77\times10^{-7} – 4{,}94\times10^{-2}$	15
Gastrointestinal disease	$1{,}77\times10^{-7} – 4{,}94\times10^{-2}$	14
Neurological disease	$1{,}59\times10^{-4} – 4{,}50\times10^{-2}$	7
Reproductive system disease	$3{,}62\times10^{-4} – 3{,}35\times10^{-2}$	12
Dermatological diseases and conditions	$5{,}14\times10^{-4} – 4{,}50\times10^{-2}$	8
Molecular and cellular functions	**p-value**	**#Molecules**
Cellular growth and proliferation	$4{,}20\times10^{-5} – 4{,}82\times10^{-2}$	14
Cellular function and maintenance	$4{,}31\times10^{-5} – 4{,}66\times10^{-2}$	14
Cell death	$1{,}30\times10^{-4} – 4{,}66\times10^{-2}$	14
Cellular movement	$1{,}40\times10^{-4} – 4{,}33\times10^{-2}$	12
Cellular development	$7{,}45\times10^{-4} – 4{,}44\times10^{-2}$	11
Canonical pathways	**p-value**	**Ratio**
VDR/RXR Activation	$7{,}66\times10^{-3}$	2/79 (0,025)
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	$1{,}18\times10^{-2}$	3/324 (0,009)
Airway Pathology in Chronic Obstructive Pulmonary Disease	$1{,}35\times10^{-2}$	1/8 (0,125)
Hepatic Fibrosis / Hepatic Stellate Cell Activation	$2{,}21\times10^{-2}$	2/142 (0,014)
Dendritic Cell Maturation	$2{,}29\times10^{-2}$	2/177 (0,011)
Transcription factors	**p-value**	**#Target molecules**
ATF3	$1{,}32\times10^{-4}$	3
ESR1	$2{,}18\times10^{-4}$	5
ATF2	$3{,}30\times10^{-4}$	3
NR1D1	$4{,}00\times10^{-4}$	2
NR3C1	$8{,}96\times10^{-4}$	6

5. Discussion

5.1. *WRAP53* mutation analysis

5.1.1. Experimental considerations

The ULL samples were sequenced for *WRAP53* mutations using the newly introduced BigDye® Direct Cycle Sequencing Kit and the BigDye® XTerminator™ Purification Kit (Applied Biosystems). The new Sanger sequencing method was successfully introduced in the lab, resulting in improved sequence quality and reduced work load. Sanger sequencing has since its development in the mid-1970s been the dominating sequencing method, and although the low throughput is a limitation compared to what next generation sequencing can provide [19], the method still has advantages in smaller targeted studies, such as the one presented. The Sanger sequencing read length is today ~750 bps, longer than in several next generation sequencing methods [19, 117]. Sanger sequencing often displays poor quality sequences the first 15–40 bases and further when exceeding ~700 bases, but no sequenced *WRAP53* fragments exceeded such lengths (table 1), and all fragments displayed high quality sequences read as close as one base from the primers.

To ensure high-quality sequencing results, all the current *WRAP53* fragments were sequenced in both forward and reverse direction. Generally, there is no guarantee that the sequencing approach used detects every single genetic alteration present, due to low percentage of mutant cells in the sample or the character of the mutation. In some cases, alterations have been observed detectable in one of the sequencing directions only. Optimally, every sequenced fragment should be interpreted in both sequencing directions, but in this pilot study performing *WRAP53* mutation screening, high-quality unidirectional sequences was considered satisfying when one direction failed. Bidirectional scoring will be of greater importance in for instance diagnostic sequencing approaches. Sequence electropherograms should generally be interpreted by at least two independent persons to ensure objective and correct sequencing data annotations. This is so far not done in this study, but the data was otherwise analyzed in accordance with the guidelines at the Department of Genetics.

The 175 ULL tumor samples included in the *WRAP53* mutation analysis are a representative selection of Norwegian breast cancer patients from the early 1990s, potentially with a negligible bias towards larger tumor sizes. Mutations are defined as

permanent genetic alterations occurring in < 1% of the population [19], but increased mutations frequencies are often observed in cancer patients. The number of ULL tumor samples may be considered slightly small in the mutation analysis coherence, but should be sufficient to detect possible medium and high frequency cancer-related *WRAP53* mutations. The size of the cohort indicates that the findings anyway should be validated in larger studies. An advantage of the ULL cohort is that it contains patient information collected over several years, making it possible to relate findings to important clinical, pathological and molecular parameters.

5.1.2. Sequence alterations in the *WRAP53* gene

No obvious somatic sequence alterations were detected in the *WRAP53* mutation analysis, indicating that mutations in the *WRAP53* gene do not seem to be a common event in breast tumorgenesis. The detected alterations were primarily single nucleotide substitutions classified as normal variations, but two likely germline indels were detected as well.

Next generation sequencing has become a revolutionizing tool in tumor characterization, and a study just published has for the first time sequenced the whole exome in 100 primary breast cancer cases to make a survey of mutated cancer genes. Somatic driver mutations were detected in frequently mutated genes like *BRCA1*, *TP53*, *RB1*, *PTEN*, *GATA3* and *PIK3CA* known to be involved in breast cancer development, confirming existing knowledge. *WRAP53* was not reported as one of the somatically altered genes [118], supporting the outcome in our study. The sensitivity of next generation sequencing is not too well known, so there is still a possibility that *WRAP53* actually might be a low-frequency mutated gene. Such mutations are however rarely conclusive factors in pathological processes, an assertion that might question the importance of *WRAP53* in breast cancer.

The R68G and A522G SNPs were the only detected non-synonymous base substitutions, causing amino acid changes from arginine and alanine to glycine, respectively. Arginine is a polar, positively charged amino acid, while alanine and glycine are non-polar and neutral. Amino acid substitutions might affect protein structure and function, but the actual functional effect is often hard to predict. As a result of the A522G SNP, the chemical properties of non-polarity and neutrality are maintained throughout the substitution, while as a result of the R68G SNP, the substitution alters the chemical properties from polar and positively charged to non-

polar and neutral. Replacing amino acids of similar chemical properties (conservative substitutions) have less effect on protein structure and function than replacing amino acids of different chemical properties (non-conservative substitutions) [19]. Because of this, the R68G SNP could be expected to affect the WRAP53 protein to a greater extent than the A522G SNP.

Sequence alterations occur frequently in the human genome, but are not necessarily drivers of malignance. Malignant alterations affect protein structure and function, and mainly occurs in functional sequences like coding regions, promoters, other regulatory sequences and splice sites [19]. The WRAP53 protein belongs to the WD40 protein family and contains six highly conserved WD40 repeats important to proper function. All the detected *WRAP53* sequence alterations are localized outside these current repeats [57]. The non-synonymous R68G and A522G alterations are in addition classified as natural sequence variants [90, 91], indicating that the detected *WRAP53* alterations do not severely affect protein function. Still, the SNPs might be involved in features of less obvious character like cancer susceptibility or therapy response modulations [119].

In 2011, Zhong et al. detected four *WRAP53* base substitutions (Phe164Leu, His376Tyr, Arg398Trp and Gly435Arg) in a group of Dyskeratosis congenita patients, and referred to the alterations as possible disease-causing factors. The DC diagnosis brings along increased cancer susceptibility, a possible link to *WRAP53* involvement in cancer development. The *WRAP53* germline mutations detected in the DC patients are located close to or in the highly conserved WD40 repeats, predicted to alter the *WRAP53* protein function. The alterations detected in the DC patients do not overlap with the *WRAP53* alterations detected in the ULL tumor samples, supporting the hypothesis that indicates benign *WRAP53* alterations. Although intriguing findings related to *WRAP53* was presented in the DC study, few included study participants is a weakness to consideration [64].

The *WRAP53* indel alterations detected in the ULL-T-142 and ULL-T-253 samples are most likely germline alterations, even though this was only confirmed in one of the patients (ULL-B-235). It cannot be excluded that the other patient (ULL-T-142) still may have a somatic mutation. Prior to the blood sample analysis, the detected indels created a hypothesis suggesting a potential functional *WRAP53* region within exon 11 targeted by somatic mutations. This hypothesis was however rejected when the ULL-T-253 indel was identified as a germline alteration. Cellular effects

induced by frameshift alterations may be hard to predict, but common consequences on the protein level are induced premature stop codons creating truncated protein unable of proper function or any function at all [19]. Premature stop codons were proven in both indel variants, relatively unexpected according to the probable germline status. Polymorphic indels tend to localize towards the end of the protein and therefore avoid the nonsense altering effects [120]. Although premature stop codon was detected in the indels, the localization towards the end of the *WRAP53* coding region might preserve protein functionality, but this remains speculative.

Another conceivable explanation to the indel alterations is that PCR and sequencing reaction artifacts due to DNA polymerase slippage in repetitive sequences (figure 13, seven following guanine bases) might cause frameshifts [19]. In our study, the *WRAP53* indel alterations were still detected by re-analyzing both tumor samples and the blood sample, indicating that artifacts probably not are the sequence altering cause. The low incidence of indel alterations requires results validation in larger studies.

The absence of somatic *WRAP53* mutations in the ULL tumor samples changed the focus from mutations towards the detected SNPs. SNP analyses are primarily not performed in tumor samples due to cancer-induced genomic aberrations, but in blood samples. This study was originally not a SNP analysis, so the tumor detected SNPs, although verified in blood, was used to investigate the *WRAP53* SNPs influence in breast cancer.

5.1.3. *WRAP53* SNPs in linkage disequilibrium

The results from the statistical analyses investigating the relation between the *WRAP53* c.–245G>C, R68G, F150F and A522G SNPs indicated a significant association between R68G, F150F and A522G. These results were further confirmed in the Haploview (v4.2, Broad Institute) analysis, defining R68G, F150F and A522G as genetic linked in a haplotype block where CCC was the most frequent genotype (figure 15).

The results verified the findings presented by Garcia-Closas et al. (2007) indicating that the *WRAP53* R68G and F150F SNPs occur in LD [65]. In addition, the A522G SNP occurred as a new marker in the same haplotype block, an observation not previously reported. High LOD scores suggested that the three SNPs are closely linked, where the greatest level of linkage was observed between the R68G and

F150F SNPs (D'=1, LOD score=23,9). The R68G and A522G occur with a D'=0,96, a value suggesting strong, but not complete linkage (D'<1). Linkage to the c.–245G>C SNP was on the other hand rejected due to low LOD scores (figure 15 and table 4). Estimates of D' values highly depends on sample sizes, so only D' values close to one are considered as reliable LD measures, indicating minimal recombination. Intermediate values are more difficult to interpret and should not be used to measure LD levels [88]. Loci inherited together as haplotypes are interesting features because the non-random heritage might reflect biological functional effects depending on specific loci and/or genotype combinations under normal or pathological conditions.

No deviation from the HW equilibrium was detected investigating *WRAP53* c.–245G>C, R68G, F150F and A522G SNPs genotype frequencies in the breast cancer patients. Cancer is a disease characterized by accumulation of genomic aberrations, and genotype specific selection promoting additional oncogenic behavior is a possible scenario in accordance with the theory of natural selection.

5.1.4. Association of *WRAP53* SNPs to clinicopathological and molecular data

The statistical analyses investigating the association of the *WRAP53* SNPs (c.–245G>C, R68G, F150F and A522G) to clinical, pathological and molecular parameters revealed no significant results except two findings related to subcellular WRAP53 protein localization. SNPs represent normal sequence variations and are generally not expected to strongly associate to clinicopathological parameters. The associations to WRAP53 localization were found intriguing by the indicative that the subcellular distribution of WRAP53 influences breast cancer prognosis (Langerod et al., unpublished data).

The exon 1β c.–245G>C heterozygous (G/C) genotype was significantly associated with nuclear WRAP53 localization, a feature of favorable breast cancer prognosis, but did not predict improved breast cancer outcome compared with the homozygous (G/G) genotype in the survival analysis (p=0,47). The low heterozygous (G/C) genotype frequency (figure 16) introduced a weakness in the statistical analyses that might affect the results. Although nuclear WRAP53 localization displays prognostic value in breast cancer, the prognostic feature may not be directly related to c.–245G>C genotype. Still, the association is intriguing and should be further validated in larger studies.

The R68G heterozygous (C/G) and minor homozygous (G/G) genotypes significantly associated with combinatory negative or positive nuclear and cytoplasmic protein localization (figure 17), the two sub-cellular protein distribution alternatives of least distinctive prognostic impact. The association with WRAP53 distribution might indicate that the c.–245G>C and R68G SNPs are involved in cellular WRAP53 localization. There are no publications discussing possible underlying causes of breast tumor differences in cellular WRAP53 localization, or probable reasons why nuclear WRAP53 localization is a positive prognostic factor (Langerod et al., unpublished data). The following perspectives will because of this remain speculative.

The c.–245G>C and R68G SNPs might be associated to protein localization, although the low number of observations slightly reduces the confidence of the current statement. The association between c.–245G>C and nuclear WRAP53 localization could be expected reflected in the breast cancer survival analyses, but it was not (p=0,47). Exon 1β is non-coding and should not be important to protein structure and function, but its potential regulatory activity could be affected by sequence alterations further involved in protein localization. Since the R68G SNP associated with the WRAP53 distributions of least utility according to prognosis, the result that indicated a statistical not significant relation between R68G and survival, although an observed trend (p=0,114), was not unexpected. There is a possibility that the R68G SNP could be located near or within a cellular localization signal, but no such *WRAP53* regions are currently known.

The increased breast cancer-specific survival by nuclear WRAP53 localization is an interesting observation. One hypothesis discuss that *WRAP53* might be involved in DNA repair mechanisms, detaining the accumulation of somatic mutations (Langerod et al., unpublished data). This is classical tumor suppressor gene functions, and a conflicting hypothesis due to previous publications discussing *WRAP53* functions in cancer. Tumor suppressor genes are in cancer frequently inactivated by mutational events [11], a feature not directly supported by the detected alterations in the *WRAP53* mutation analysis, but loss of heterozygosity and epigenetic methylation are as well mechanisms that might affect *WRAP53* function.

Another possible hypothesis concerning sub-cellular WRAP53 protein localization involves the *WRAP53* overexpression observed in cancer cells. *WRAP53* overexpression increases the p53 antisense transcript levels, subsequently increasing

cellular p53 activity [1]. p53 is a key molecule in DNA repair and apoptosis [40], important anti-tumorigenic mechanisms. *WRAP53* is as well reported to regulate cellular levels of mutated p53 [1], an intriguing observation since *TP53* mutated tumors with nuclear WRAP53 localization in particular display improved prognosis (Langerod et al., unpublished data).

5.1.5. *WRAP53* SNPs and survival analyses

One of the four SNPs analyzed, *WRAP53* F150F, was found significantly associated with breast cancer-specific survival. It is uncommon that one SNP displays such significant influence on survival, and in particular when the alteration is silent. Despite less significant results, a clear trend was observed in the R68G (p=0,114) and A522G (p=0,133) SNPs, indicating that the respective heterozygous and minor homozygous genotypes overall associated with a poorer breast cancer prognosis.

The genetically linked *WRAP53* SNPs (R68G, F150F and A522G) and the defined haplotype block make it difficult to determine the actual cause of the reduced survival in F150F heterozygous (C/T) and minor homozygous (T/T) genotype carriers compared to the major homozygous (C/C) genotype carriers. The difference in survival might be a direct cause of the F150F SNP, or it could be an indirect effect caused by the *WRAP53* haplotype. The publication by Garcia-Closas et al. (2007) indicates that the LD R68G and F150F SNPs also are genetically linked to *TP53* SNPs [65]. *TP53* is a well-known prognostic marker of breast cancer and some SNPs, e.g. Arg72Pro, have been suggested to associate with survival [121]. There is a possibility that the haplotype involves other SNPs of greater influence to breast cancer survival, mediating the difference in survival observed by the *WRAP53* F150F SNP. An extended SNP analysis including *WRAP53* and neighboring genes will be necessary to outline this possibility.

The survival analyses stratified for *TP53* mutation and ER status revealed interesting results regarding *WRAP53* SNPs (R68G, F150F and A522G) genotypes. The heterozygous and minor homozygous genotypes were significant (R68G and A552G) or borderline significant (F150F) associated with reduced survival in wt *TP53* tumors, and overall associated with significant reduced survival in ER positive tumors compared to the major homozygous genotypes. No genotype-dependent survival differences were observed in the *TP53* mutated and ER negative tumors.

TP53 mutation status and ER status are strongly associated with different breast cancer subtypes. *TP53* mutations and ER negativity occur commonly together and is related to poor prognosis compared to patients with wt *TP53* and ER positivity [30]. The overall poor prognosis might be an explanation why no *WRAP53* SNPs genotype-dependent differences in survival was detected in the *TP53* mutated and ER negative tumors. The alternative hypothesis is that the SNPs genotype-related survival only affects the wt *TP53* and/or ER positive tumors.

ER positivity is a tumor characteristic that is associated with favorable breast cancer prognosis. This is among other things due to available targeted hormonal therapies like tamoxifen and aromatase inhibitors. The *WRAP53* SNPs were able to split the patients into a good and poor prognosis group, and may have the potential to predict who benefits from hormonal treatment. Further investigation is needed to validate the findings and understand the possible mechanisms. An analogue to the observation in the ER positive tumors is the metabolic capacity of Cytochrome P450 2D6 **(**CYP2D6) and the response to tamoxifen treatment. Tamoxifen is an estrogen receptor antagonist prodrug metabolized by CYP2D6, a gene of highly phenotypic variabilities that affects the enzymatic capacity. Poor metabolizers have very little or no CYP2D6 activity and responds poorly to tamoxifen treatment while responders display normal metabolic capacity and responds well [122], roughly dividing ER positive tumors in two groups.

The underlying causes of the different prognostic effects seen in wt *TP53* versus mutated *TP53*, and in ER positive versus ER negative tumors are hard to predict. Since the *WRAP53* SNPs (R68G, F150F and A522G) induces similar effects on survival and the SNPs are inherited together and possibly selected for, *WRAP53* seem to be a gene of importance in breast tumorgenesis. The significant differences in survival observed in wt *TP53* and in ER positive tumors might be used therapeutically to avoid over-treatment of the *WRAP53* SNPs major homozygous genotype carriers, and perhaps intensify the treatment of heterozygous and minor homozygous genotype carriers.

The *WRAP53* sequence alterations detected in the current study are interesting in relation to breast tumorgenesis, and the findings should be validated in larger studies. The potential of *WRAP53* SNPs used as biomarkers in wt *TP53* and ER positive breast cancer patients is intriguing, and whether this may be related to hormonal treatment responses should be investigated. Further research is as well

needed to in general investigate the importance of *WRAP53* in breast cancer, due to the limited knowledge about *WRAP53* functionality. The results from this study indicate that *WRAP53* might be a marker of prognostic value in breast cancer, and the importance of the *WRAP53* SNPs genotypes in relation to protein localization and breast cancer-specific survival should be further studied.

5.2. Gene expression study

5.2.1. Experimental considerations

Breast cancer cell lines are model systems extensively used to investigate and better understand breast cancer. In this study, MCF-7 and MDA-MB-231 were used to explore the cellular responses to *WRAP53* knockdown. MCF-7 and MDA-MB-231 are the most commonly used cell lines in breast cancer research, and were chosen because of their extensive use and their differences according to e.g. *TP53* mutation and ER status, features characteristic of the two main sub-types of breast cancer, the luminal and basal-like. Differences in *TP53* mutations status was especially emphasized in the experimental study design due to the reported *WRAP53*-mediated p53 regulatory mechanism [1]. Other cell lines could have been included in the study as well, but due to the probable hypothesis creating study content, uncertainties according to expected findings and the economical aspect, the size of the study was moderated.

The MCF-7 and MDA-MB-231 *WRAP53* knockdown was performed using an siRNA targeting *WRAP53* exon 2 (Qiagen) inducing an isoform-independent knockdown. Another *WRAP53* siRNA targeting exon 6 is offered by the same vendor, but the *WRAP53* exon 2 targeting siRNA has proven to be most efficient according to experiences made by Farnebo and colleagues at CCK (90% WRAP53 mRNA knockdown) [123]. *WRAP53α* and *WRAP53γ* generates certain splice variants lacking exon 2, the siWRAP53#2-specific binding site (figure 7), which is a possible cause of the incomplete *WRAP53* knockdown in our system. In this study, only one siRNA was used in the *WRAP53* knockdown to restrain the number of variables affecting the cellular responses to *WRAP53* depletion. A mix containing several siRNAs targeting the same genes is commonly used in transfection experiments.

RNA was harvested from the MCF-7 and MDA-MB-231 cells 40 and 72 hours after siRNA transfection. *WRAP53* depletion is previously reported to induce apoptosis in cancerous cells 48–62 hours after transfection [3]. This was not observed when the MCF-7 and MDA-MB-231 cells were siC or siWRAP53#2 transfected and subsequently stained for Annexin V at 36, 48, 60 and 72 hours after transfection. The observations from the optimization of the siRNA transfections uncovered that the *WRAP53* depleted MCF-7 and MDA-MB-231 cells did not enter apoptosis 72 hours after transfection, which made it possible to obtain a more complete *WRAP53* knockdown.

RNA was harvested from the MCF-7 and MDA-MB-231 siC and siWRAP53#2 transfected cells cultured in triplicates, and from the NTC cells grown as single samples (table 2). Microarray-based analyses are multi-step procedures influenced by numerous sources of variation, but analyzing samples in replicates increases the results reliability. To compensate the lack of NTC cell replicates, the samples were analyzed twice using microarrays as technical duplicates. By using biological replicates, experimental variation was measured and made it possible to remove outliers from the dataset. The replicates contributed to determine if gene expression differences between measurements were caused by actual biological variations or randomly, and increased robustness of the conclusions. High cost of microarrays often restrains the number of included replicates, but the costs should always be considered versus data quality. Technical replicates are used in larger sample series to control that the experimental conditions are reproducible and do not affect the gene expression experiments and results over time [124].

Microarray data quality control was performed using the GeneSpring GX 12.0 software (Agilent Technologies) PCA plot to capture dataset variance. Samples from the same experimental condition are expected to group closely to each other [104], and deviant sample grouping might be explained by poor sample quality or actual biological variation. By eliminating one sample outlier (MDA-MB-231 siWRAP53#2 transfected cells, 40 hours, triplicate 1), the cell line-specific samples grouped together in the PCA plot (figure 22). There was on the other hand no specific trends in the alignment of each respective group according to siRNA treatment (NTC cells, siC and siWRAP53#2 transfected cells). This may be due to the small number of samples analyzed or experimental variation, but the most probable cause is that the actual siWRAP53#2-induced gene expression changes are minor to and dominated by the

cell line-specific gene expression patterns. To minimize the impact of experimental variabilities, the samples were placed on the microarrays randomly.

5.2.2. Considerations regarding bioinformatics and statistical analyses

Microarray gene expression analyses generate huge amounts of data and introduce challenging assignments according to data processing and interpretation.

In accordance with the recommendations from Agilent Technologies, the microarray data was normalized using the 75th percentile shift method. Percentile normalization presumes that a certain level of expression values is equal for all arrays. Both the 50th and 75th percentile is frequently used, as well as other methods like quantile and housekeeping genes normalization [104, 125]. However, in gene expression analyses, the 75th percentile normalization is often preferred because it is a more robust method for measuring small intensity values, and because the percentile should be well within the range of detectable data [126].

The statistical significant *WRAP53* knockdown-induced gene expression changes were identified using SAM. SAM is a commonly used analytical tool in microarray experiments, handling huge datasets in a non-parametric approach that not requires normal distributed data [127]. Compared to conventional t-tests for the same approach, SAM is proven superior for microarray data analyses [106]. Several statistical correction methods have been developed to restrain the fraction of false positive results by analyzing huge amounts of data. Bonferroni is a frequently used method, but the analysis is very conservative. FDR is a more sensitive method, an important consideration choosing SAM in this study, even though the risk of false positive results increases coincidentally compared to the Bonferroni method [106].

IPA was used in the pathway analysis to associate significant gene expression alterations to biological functions. A weakness of the analysis is that the outcomes are merely based on the dataset gene identifications, excluding the genes significance levels according to expression alterations. A more weighted analysis including such information could to a greater extent list the most significant biological results in the datasets. Another challenge using IPA is the wideness of the biological functional outcomes, an aspect that might complicate the work of understanding connections and draw conclusions. An advantage using IPA is that that the Ingenuity Knowledge Base contains manually curated data from scientific articles [108].

5.2.3. *WRAP53* knockdown efficiency and RNA isolation

The MCF-7 and MDA-MB-231 *WRAP53* knockdown efficiency was evaluated by qPCR and Western blot analyses, and by qPCR quantified to 75% and 70% (figure 21), respectively. Despite incomplete knockdown, the results were considered satisfactory. The loading control (here β-actin) was used to ensure proper Western blot interpretations. The controls are used to assure that gel lanes are equally loaded with sample, an important aspect when protein expression levels are compared between different samples. Loading controls usually display constant cell type-specific expressions, so the expression levels should not vary between sample lanes [128], indicating actual siRNA-induced *WRAP53* depletion in this study (figure 20).

The *WRAP53* knockdown was more prominent in the MCF-7 than the MDA-MB-231 cell line reflected by lower protein band intensities (figure 20). This might be a result of the initial abundant MDA-MB-231 *WRAP53* expression levels or a delayed knockdown compared to the MCF-7 cell line. Another feature is that MCF-7 cells do not express the *WRAP53γ* isoform [1], but this is probably not the explanation to the differences in *WRAP53* knockdown efficiency, since *WRAP53γ* overall is the least abundant expressed isoform [78].

The purity of the MCF-7 and MDA-MB-231 RNA samples were measured using the NanoDrop® ND-1000 Spectrophotometer. The calculated 260/280 ratios from these measurements were slightly lower (ranging from 1,75–1,95) than the recommendations for pure RNA. Reduced ratios indicate presence of co-purified contaminants like proteins or phenols absorbing at or near 280 nm [75]. The TRIzol® reagent used for RNA isolation is a phenolic solution containing guanidine isothiocyanate strongly absorbing at 230 and 270 nm [129], and residues might affect RNA purity. The measured 260/230 RNA ratios were primarily within the recommended range of pure RNA.

Despite the low 260/280 RNA ratios, there were no signs that reduced RNA purity affected the quality of the microarray analyses. The yield of cRNA and the specific activity according to cRNA Cy3 incorporation (Appendix C, step 4) fulfilled the quality requirements in the microarray procedure. Sample quality was after microarray scanning further evaluated by the Feature Extraction software (Agilent Technologies), approving all samples.

5.2.4. Significance Analysis of Microarrays

Cell line specific gene expression changes as a response to siRNA *WRAP53* knockdown was investigated using SAM. *WRAP53* depletion primarily resulted in upregulation of gene expression, most clearly detected 72 hours after transfection. No significant changes were detected comparing gene expression patterns in the NTC cells and the siC transfected cell within each cell line, indicating that the siRNA transfection procedure itself did not influence with the cellular gene expression levels. Gene expression alterations in the siWRAP53#2 transfected cells were thus caused by *WRAP53*-targeted gene silencing. In addition, *WRAP53* was observed significantly downregulated in both cell lines 72 hours after transfection, designating efficient knockdown.

The results generated from SAM reflected the qPCR and Western blot findings, indicating that the *WRAP53* knockdown response occurred more slowly in the MDA-MB-231 than the MCF-7 cell line. The MCF-7 gene expression alterations occurred more prominently throughout the whole experiment, with overall greater fold change and lower FDR values than observed in the MDA-MD-231 cell line (Appendix E, table 11 and 12). The aim was to generate SAM gene lists of manageable sizes (300–600 genes) for the downstream data analyses, and with low FDR values (ideally FDR < 5%) to obtain as reliable gene datasets as possible restraining the incidence of falsely positive results. In the MDA-MB-231 cell line, data reliability according to significant differentially expressed genes was reduced, giving an increased FDR value. In this case, satisfactory FDR values resulted in fewer listed genes, a disadvantage in the following pathway analyses according to the aim of widely studying the cellular responses by *WRAP53* knockdown. To obtain a list of MCF-7 comparable numbers of genes a FDR value of 10,5% was accepted, an action giving decreased but still sufficient data reliability, but an important aspect to consider in further data analyses and interpretations.

5.2.5. Pathway analyses

To investigate the biological functions associated with MCF-7 and MDA-MB-231 upregulated genes 72 hours after *WRAP53* siRNA transfection, pathway analyses using IPA were performed. The IPA calculated p-values helped identify significant cellular functions, pathways and molecules in the *WRAP53* siRNA-induced gene expression alterations, and acted as starting points for further investigation to understand the biological implications of the significant results. In the process

towards biological comprehension, the necessity of exploring supportive evidence and potential interesting biological results, even without statistical significance, should not be underestimated [109].

Overall, the most interesting findings in the pathway analyses were the results significantly associating cancer and cancer-related characteristics with the MCF-7 and MDA-MB-231 upregulated gene datasets. In both cell lines, cancer was listed as the most significant disease (table 6 and 7), and the same was observed analyzing the MCF-7 and MDA-MB-231 mutually upregulated genes (table 9). Both cell lines, with an extensive number of mapped molecules, displayed significant associations to the biological functions of cellular movement, growth and proliferation, well acknowledged cancer hallmarks [18]. This may indicate that *WRAP53* knockdown actually induces cellular responses favoring tumorigenic activities. The pathway analyses results are merely based on the gene expression differences induced by *WRAP53* depletion, and cancer associations were no obvious outcomes. In accordance with the aims of this study, the cancer-related results will be the main focus of the discussion.

In the MCF-7 cell line, *WRAP53* knockdown did not result in any statistical significant canonical pathways ($p > 0,05$) (table 6). However, the p53 signaling still occurred as an interesting result ($p=0,097$). *WRAP53* is reported to regulate endogenous p53 mRNA and protein levels, and downregulated *WRAP53* expression to suppress p53 induction upon DNA-damage [1]. Increased p53 activity was thus not an expected response to *WRAP53* knockdown, but the p53 transcription factor regulates the expression of a myriad of target genes which might cause increased p53 signaling. The MCF-7 upregulated genes in the p53 signaling pathway primarily encoded cell cycle inhibitors and p53-induced pro-apoptotic proteins, features of possible relation to the occurrence of cell death as the top significant biological function. *WRAP53* depletion by siRNA treatment is reported to induce apoptosis in cancerous cells 48–62 hours after transfection [3], but this was not observed in either the MCF-7 or MDA-MB-231 cell line. The upregulated p53 signaling might still be a post-transfectional cellular stress response, a hypothesis supported by the upregulation of cell death-related genes counteracting with the outcomes of cancer promoting characteristics.

Another pathway of notice from the analysis of MCF-7, although not statistical significant, was the Vitamin D Receptor (VDR)/Retinoic Acid X Receptor

(RXR) activation, a significant pathway when analyzing the mutually upregulated MCF-7 and MDA-MB-231 genes (table 9). The VDR is a nuclear receptor that transcriptionally regulates its target genes by binding the vitamin D ligand. Activated VDR dimerizes with the RXR, which modulates the transcriptional activity [130]. The VDR receptor are under normal conditions involved in multiple cellular pathways, but different polymorphic variants are reported involved in several types of cancers as well [131], a possible relation to the cancer-related outcomes. Despite interesting biological results, the MCF-7 borderline significant p-values and relatively low ratios require careful interpretation.

Of the top five listed significant transcription factors, *TP53* and *TP63* were found to regulate a considerable fraction of the MCF-7 upregulated genes (table 6). p53 dysregulation is a well-known oncogenic feature, but *TP53* also regulates numerous target genes under normal cellular conditions. *TP63* belongs to the "p53 family" of genes and displays a *TP53* sequence homology in the DNA-binding domain of > 60%. The sequence homology results in mutually shared target genes, and may explain the observations in the MCF-7 dataset even though *TP63* primarily is involved different cellular processes than *TP53*, such as skin and limb formation [132]. MCF-7 *WRAP53* depletion seems to upregulate numerous *TP53* and *TP63* regulated genes, an outcome presumed to be associated with a possible cellular stress response. The *ATF3* and *FOSL2* were also interesting transcription factors, considering the cancer-related results from the pathway analysis. *ATF3* is activated by various signals, including many encountered by tumor cells, but is also involved in cellular stress responses [133], while *FOSL2* belongs to a gene family important in cell proliferation, differentiation, and transformation regulation [134].

Although cancer was listed as the most significant disease, and cellular movement, growth and proliferation as top biological functions, the most striking result from the MDA-MB-231 pathway analysis was that most categories displayed immune systemic characteristics (table 7). Biological functional results including inflammatory response, antigen presentation, cell-to-cell signaling and interactions, and cellular development are all immune-related mechanisms. Three of the listed canonical pathways were also immune response-associated and identified interleukin-17A (IL-17A) as a molecule of particular interest. IL-17A belongs to the interleukin 17 (IL-17) family of proinflammatory cytokines, and induces the production of other cytokines and chemokines creating inflammatory environments. The IL-17 family is

associated with a diversity of human diseases [135], but the role of IL-17 in malignancies is unclear. Conflicting data exists concerning a potential involvement in cancer, indicating both IL-17-induced angiogenesis and T-cell-mediated tumor rejection [136]. In addition, one study has reported an association between IL-17A gene polymorphisms and breast cancer risk and prognosis [137]. Tumor-promoting inflammations have the last decade become a field of great attention in research on tumor development, where innate immune cells are presumed to create an inflammatory tumor environment promoting multiple cancer hallmarks capabilities [18].

The IPA results emphasizing the immune system were further reflected in the transcription factors that regulated the expression levels of the MDA-MB-231 upregulated genes (table 7). NFκB is a complex of gene regulatory proteins, including *RELA*, activated in many stressful, inflammatory and innate immune responses. In the MDA-MB-231 gene set, NFκB and *RELA* were the transcription factors regulating the expression of the major fraction of upregulated genes. Active NFκB increases the transcription level of hundreds of target genes involved in inflammatory and innate immune responses, but excessive NFκB signaling is also observed various human cancers [5].

Another interesting transcription factor was the estrogen receptor 1 (*ESR1*), mediating the actions of estrogen. Normally, *ESR1* is involved in sexual development and reproductive function, but is also detected in pathological states like endometrial and breast cancer [138]. *ERS1* is as well a key molecule in breast cancer by characterization of the main subtypes [30], and as a predictive marker of targeted hormonal therapy. *ESR1* is one of the first evidences of personalized medicine based on molecular knowledge [23]. It is therefore intriguing to find a possible association to *WRAP53*.

The upregulation of immune responses in the MDA-MB-231cell line most likely was a response to *WRAP53* depletion, although there are to our knowledge no publications that indicate such an association. Gene expression silencing through RNA interference using siRNAs has become a powerful tool to study gene functions. This strategy relies on a high degree of specificity to obtain efficient gene silencing, but different non-specific effects, including activation of the immune system, have been reported. The pathways for siRNA recognition and immune system activation are not completely understood, but Toll-like receptors (TLRs) seem to be central

participants. TLRs activate the innate immune system by recognizing pathogen-associated molecular patterns like viral dsRNAs, the same molecular structure found in siRNAs. TLR ligand activation by siRNAs triggers downstream activation of the interferon regulatory factor, NFκB and MAPK pathways, leading to increased expression levels of interferon (INF) and proinflammatory cytokines mediating an unwanted immune response [139, 140]. The immunogenic activity of siRNAs mainly seem to be associated to specific structural motifs, delivery methods using cationic lipids and increasing siRNA concentrations, although contradictory observations exist [140]. To avoid siRNA-induced non-specific immune responses it is recommended to be conscious of the siRNA design, use low siRNA concentrations (10–20 nM) and several siRNAs targeting a specific gene [141].

In the MDA-MB-231 cell line, there are reasons to assume that the immune response outcomes from the pathway analyses were actual effects of the *WRAP53* knockdown. The responses were not present in the MDA-MB-231 siC transfected cells, indicating satisfactory transfection conditions. No typical INF stimulated genes (including *JAK1, TYK2, STAT1, STAT2* and *IRF9*) [139] were upregulated by the siWRAP53#2 transfection. However, it is not possible to totally exclude the alternative explanation of a non-specific immune response. The results should therefore be validated, e.g. by using other siRNAs to target *WRAP53* in the MDA-MB-231 cell line [141].

Since the MCF-7 and MDA-MB-231 pathway analyses displayed similar results including cancer and cancer-related characteristics, the aim was to find indications of one or several biological functions associated with *WRAP53* knockdown prominent enough to be detected in both cell lines. According to the general cancer features like cellular growth, proliferation and movement, there is a possibility that *WRAP53* could be involved in cancer-related biological functions present in different cells types and tissues. A Venn-diagram tool [115] identified 29 shared MCF-7 and MDA-MB-231 upregulated genes (figure 24, table 8), and the subsequent pathway analysis results listed cancer as the most significant disease (table 9). Cellular growth, proliferation and movement were repeatedly displayed as significant biological functions, supporting the assumption that *WRAP53* might be involved in cancer promoting activities. The MCF-7 and MDA-MB-231 genetic overlap was not extensive (29 out of 299 and 249 genes, respectively), but cancer still emerged as a significant outcome. The list of mutually upregulated genes (table 8)

was investigated pursuant to classical cancer-associated features like proliferation, cell cycle control, DNA repair, apoptosis and telomerase function using the NCBI Gene database[12], but no distinctive functions emerged. Cancer is a heterogeneous group of genetic diseases involved in dysregulation of complex cellular pathways and mechanisms. Lack of consistent functions of the mutually expressed MCF-7 and MDA-MB-231 genes made it difficult to further specify the cancer-related findings.

The pathway analysis of the mutually upregulated genes displayed results comparable with the cell line-specific analyses, but there were some deviant outcomes. VDR/RXR activation was listed as the most significant canonical pathway, and *NR1D1* and *NR3C1* were listed as significant transcription factors in addition to the previous reported *ATF3*, *ESR1* and *ATF2* (table 9). IPA assess the biological functions most significant to the dataset of interest [108], but the number of genes analyzed might affect the levels of statistical significance. In the MCF-7 pathway analysis, VDR/RXR activation was not displayed as a statistical significant result although the ratio (table 6) was higher compared to the corresponding significant outcome of VDR/RXR activation by analysis of the 29 MCF-7 and MDA-MB-231 mutually upregulated genes (table 9). Small number of genes may influence the outcome of statistical analyses, and therefore, the importance of VDR/RXR activation by *WRAP53* knockdown should be interpreted cautiously. The *NR1D1* and *NR3C1* transcription factors showed statistical significance in the separate MCF-7 and MDA-MB-231 analyses as well, although not as the top five most significant results. This strengthens the data reliability despite the low gene number in the MCF-7 and MDA-MB-231 mutual analysis. *NR1D1* and *NR3C1* are receptors involved in cellular proliferation, differentiation and inflammatory responses [142, 143], relevant mechanisms in cancer.

The MCF-7 and MDA-MB-231 *WRAP53* knockdown resulted in cell type-specific biological responses, but cancer and cancer-related characteristics like cellular growth, proliferation and movement emerged as common denominators. The MCF-7 and MDA-MB-231 are quite different cell lines, a possible explanation to the differences in cellular responses from silencing the same gene. The *WRAP53* knockdown responses might as well reflect the breast cancer subtypes from which the cell lines were retrieved, and give knowledge of their different biology. Although

[12] Available from http://www.ncbi.nlm.nih.gov/gene/

TP53 mutation status was an important consideration in the study design, it is difficult to directly relate the cell line differences of *TP53* mutation and ER status to the pathway analyses outcomes. In the MCF-7 analysis, p53 signaling and the *TP53* and *TP63* transcription factors emerged as interesting results. The same results were not observed in the MDA-MB-231 cell line, a difference that might be related to *TP53* mutation status. The MCF-7 cell line is wt *TP53*, and there is a possibility that the *TP53* mutated MDA-MB-231 is unable to activate p53 signaling. In addition, the statistical significance of the *ESR1* transcription factor was observed in the MDA-MB-231 cell line, although the cell line is derived from an ER negative tumor. This indicates that *WRAP53* knockdown might induce different responses in distinct cell types and tissues, making it difficult to study the specific biological aspects of interest. The possible non-specific immune systemic effects in the MDA-MB-231 cell line might as well be a contributory factor to the differences observed.

WRAP53 depletion has been reported associated with apoptosis in cancerous cells and progressive telomere shortening [3, 64], but cancer-related outcomes were not the expected and evident results by *WRAP53* knockdown. The findings support the hypothesis of *WRAP53* involvement in cancer development, but IPA relates the gene datasets to relatively wide biological functions and makes it hard to answer specific questions. No specific pathway promoting an explanation of *WRAP53* influence in breast cancer was identified, but the results do not exclude *WRAP53* as a gene involved in tumorgenesis. The results should be validated in patient cohorts or by functional studies including other *WRAP53* siRNAs and breast cancer cells lines to ensure efficient, specific knockdown. Further research will be needed to understand the possible influence of *WRAP53* in breast cancer, and the potential of *WRAP53* as a therapeutic target. Increased knowledge concerning normal *WRAP53* function should also be pursued prior to a future potential in targeted cancer therapy.

6. Conclusion

The significance of *WRAP53* in breast cancer was investigated performing a *WRAP53* mutation analysis in primary breast carcinomas and a gene expression study in *WRAP53* depleted breast cancer cell lines as an approach to increase the knowledge about *WRAP53* cellular signaling pathways and networks.

Somatic *WRAP53* mutations do not seem to be important events in breast tumorgenesis, but detected *WRAP53* SNPs, directly or indirectly, displayed significant impact on breast cancer-specific survival. The haplotype SNPs, R68G, F150F and A522G, all displayed genotype-dependent survival, and the effect was most prominent when stratifying for the *TP53* mutation and ER status. The survival of wt *TP53* and ER positive tumors were influenced by *WRAP53* SNPs genotypes, while the outcomes in *TP53* mutated and ER negative tumors were unaffected. The exon 1β c.–245G>C SNP was found associated with nuclear WRAP53 localization, a feature of favorable prognostic impact in breast cancer (Langerod et al, unpublished data). The SNP results indicate that *WRAP53* might be a marker of prognostic value in breast cancer. The SNP haplotype however complicates the interpretation of *WRAP53* significance in breast tumorgenesis, since the R68G SNP occurs in a haplotype that includes at least one SNP known to affect breast cancer prognosis [65, 121].

WRAP53 depletion in the MCF-7 and MDA-MB-231 cell lines induced gene expression alterations and cell type-specific responses, but cancer and cancer-related characteristics emerged as common denominators. These results suggest that *WRAP53* might be involved in tumorgenesis, although no classical cancer genes were identified among the genes that showed expression alterations. No obvious cancer-related pathways or networks were identified to clarify the actual roles of *WRAP53* in cancer. Further research will thus be needed in the search for new therapeutic targets related to *WRAP53* function, and increased knowledge about normal *WRAP53* functionality should also be pursued prior to therapeutic targeting.

The current study aimed to elucidate the function of *WRAP53* using two different approaches. Although is difficult to directly connect the results from the mutation analysis and gene expression study, they both suggest that *WRAP53* might be involved in tumorgenesis. The results should be validated in independent studies, and further research will be needed to understand the importance of *WRAP53* in breast cancer.

7. Future aspects

Further investigation of the *WRAP53* gene is of current interest according to its potential involvement in tumorgenesis. The association with p53 regulation and telomerase activity are also features of interest in this context. The exact biological functions of *WRAP53* are today probably not fully known, but will be important to reveal in the process towards a greater understanding of *WRAP53* significance in cancer development, and especially in breast cancer that was the focus of this study. Suggestions to further *WRAP53* studies are subsequently listed;

- DNA sequence the *WRAP53* start exon 1α. Exon 1α is the p53 regulatory region of *WRAP53*, and sequence alterations might influence the sense/antisense regulation and affect p53 activity.
- Validate the findings from the *WRAP53* mutations analysis in larger studies, especially to investigate the frequency and possible importance of indel alterations.
- Correlate the *WRAP53* SNPs (R68G, F150F and A522G) and haplotypes with *WRAP53* gene expression levels, and further relate these results to breast cancer-specific survival.
- Investigate whether the *WRAP53* SNPs are involved in sub-cellular WRAP53 protein localization due to the prognostic impact observed in breast cancer.
- Try to modulate the effect induced by *WRAP53* SNPs on protein function. Restricted knowledge about WRAP53 structure may thus be an analysis limitation.
- Further investigate the data obtained from the gene expression study using other bioinformatics tools, including gene ontology approaches in order to study *WRAP53* functionality.
- Repeat the gene expression study using other *WRAP53* siRNAs to validate the findings, but also investigate if immune response activation in the MDA-MB-231 is a non-specific response or actually caused by *WRAP53* depletion.
- Irradiate cultured cell lines and compare *WRAP53* gene expression patterns before and after radiation. Radiation is a stress stimulus causing DNA damage responses, and may be an approach to investigate if *WRAP53* is involved in DNA repair mechanisms.

8. Reference list

1. Mahmoudi, S., et al., *Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage.* Molecular cell, 2009. **33**(4): p. 462-71.
2. Venteicher, A.S., et al., *A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis.* Science, 2009. **323**(5914): p. 644-8.
3. Mahmoudi, S., et al., *WRAP53 promotes cancer cell survival and is a potential target for cancer cell therapy.* Cell Death and Disease, 2011.
4. Norwegian Cancer Society. *Hva er kreft? Fakta om kreft og oppdaterte tall på kreft i Norge*. 14.03.2011 29.11.2011]; Available from: http://www.kreftforeningen.no/om_kreft/hva_er_kreft.
5. Alberts, B., et al., *Molecular Biology of the Cell*. 5 ed2008: Garland Science.
6. Jemal, A., et al., *Global cancer statistics.* CA: a cancer journal for clinicians, 2011. **61**(2): p. 69-90.
7. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* International journal of cancer 2010.
8. Norwegian Cancer Society. *Forebygg kreft - skift livsstil*. 09.10.2006 [cited 2011 30.11.2011]; Available from: http://www.kreftforeningen.no/aktuelt/aktuelt_i_media/forebygg_kreft___skift_livsstil_5991.
9. Haldorsen, T., et al. *Cancer in Norway 2009 - Cancer incidence, mortality, survival and prevalence in Norway*. 2011; Available from: http://www.kreftregisteret.no/Global/Publikasjoner%20og%20rapporter/Cancer%20in%20Norway/Cancer_in_Norway_2009_and_Special_Issue.pdf.
10. Manchester, K.L., *Theodor Boveri and the origin of malignant tumours.* Trends in cell biology, 1995. **5**(10): p. 384-7.
11. Balmain, A., J. Gray, and B. Ponder, *The genetics and genomics of cancer.* Nature genetics, 2003. **33 Suppl**: p. 238-44.
12. Norwegian Cancer Society. *Arvelig kreft.* 30.11.2011]; Available from: http://www.kreftforeningen.no/om_kreft/arvelig_kreft_8096.
13. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.
14. Lyons, J.G., et al., *Clonal diversity in carcinomas: its implications for tumour progression and the contribution made to it by epithelial-mesenchymal transitions.* Clinical & experimental metastasis, 2008. **25**(6): p. 665-77.
15. Axelrod, R., D.E. Axelrod, and K.J. Pienta, *Evolution of cooperation among tumor cells.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(36): p. 13474-9.
16. Jordan, C.T., M.L. Guzman, and M. Noble, *Cancer stem cells.* The New England journal of medicine, 2006. **355**(12): p. 1253-61.
17. Norton, L. and J. Massague, *Is cancer a disease of self-seeding?* Nature medicine, 2006. **12**(8): p. 875-8.
18. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
19. Strachan, T. and A. Read, *Human Molecular Genetics*. 4 ed2011: Garland Science.
20. Croce, C.M., *Oncogenes and cancer.* The New England journal of medicine, 2008. **358**(5): p. 502-11.
21. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
22. Boyd, L.K., M. Xueying, and Y.-J. Lu, *Use of SNPs in cancer predisposition analysis, diagnosis and prognosis: tools and prospects.* Med. Diagn. 3(3):313-326, 2009.
23. DeVita, Hellman, and Rosenberg, *CANCER: Principles & Practice of Oncology*. 8 ed. Vol. 2. 2008: Lippincott Williams & Wilkins.
24. Polyak, K., *Breast cancer: origins and evolution.* The Journal of clinical investigation, 2007. **117**(11): p. 3155-63.
25. Polyak, K., *Breast cancer gene discovery.* Expert reviews in molecular medicine, 2002. **4**(18): p. 1-18.
26. Gartner, L.P. and J.L. Hiatt, *Color textbook of histologi*. 3 ed2007: Saunders Elsevier. 485-488.
27. Bombonati, A. and D.C. Sgroi, *The molecular pathology of breast cancer progression.* The Journal of pathology, 2011. **223**(2): p. 307-17.
28. Allred, D.C., S.K. Mohsin, and S.A. Fuqua, *Histological and biological evolution of human premalignant breast disease.* Endocrine-related cancer, 2001. **8**(1): p. 47-61.

29. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
30. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(19): p. 10869-74.
31. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(14): p. 8418-23.
32. Perou, C.M. and A.L. Borresen-Dale, *Systems biology and genomics of breast cancer.* Cold Spring Harbor perspectives in biology, 2011. **3**(2).
33. Herschkowitz, J.I., et al., *Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors.* Genome biology, 2007. **8**(5): p. R76.
34. Langerod, A., et al., *TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer.* Breast cancer research : BCR, 2007. **9**(3): p. R30.
35. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012.
36. Linzer, D.I.H. and A.J. Levine, *Characterization of a 54K Dalton Cellular SV40 Tumor Antigen Present in SV40-Transformed Cells and Uninfected Embryonal Carcinoma Cells.* Cell, 1979. **17**.
37. Lane, D.P. and L.V. Crawford, *T antigen is bound to a host protein in SV40-transformed cells.* Nature, 1979. **278**.
38. Eliyahu, D., et al., *Participation of p53 cellular tumour antigen in transformation of normal embryonic cells.* Nature, 1984. **312**(5995): p. 646-9.
39. Baker, S.J., et al., *Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas.* Science, 1989. **244**(4901): p. 217-21.
40. Levine, A.J. and M. Oren, *The first 30 years of p53: growing ever more complex.* Nature reviews. Cancer, 2009. **9**(10): p. 749-58.
41. Olivier, M., M. Hollstein, and P. Hainaut, *TP53 mutations in human cancers: origins, consequences, and clinical use.* Cold Spring Harbor perspectives in biology, 2010. **2**(1): p. a001008.
42. Levine, A.J., W. Hu, and Z. Feng, *The P53 pathway: what questions remain to be explored?* Cell death and differentiation, 2006. **13**(6): p. 1027-36.
43. Varna, M., et al., *TP53 status and response to treatment in breast cancers.* Journal of biomedicine & biotechnology, 2011. **2011**: p. 284584.
44. Borresen-Dale, A.L., *TP53 and breast cancer.* Human mutation, 2003. **21**(3): p. 292-300.
45. Baynes, C., et al., *Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk.* Breast cancer research : BCR, 2007. **9**(2): p. R27.
46. Langerod, A., et al., *The TP53 codon 72 polymorphism may affect the function of TP53 mutations in breast carcinomas but not in colorectal carcinomas.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2002. **11**(12): p. 1684-8.
47. Pharoah, P.D., N.E. Day, and C. Caldas, *Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis.* British journal of cancer, 1999. **80**(12): p. 1968-73.
48. Olivier, M., et al., *The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(4): p. 1157-67.
49. Takahashi, S., et al., *Prediction of breast cancer prognosis by gene expression profile of TP53 status.* Cancer science, 2008. **99**(2): p. 324-32.
50. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(38): p. 13550-5.
51. Tycowski, K.T., et al., *A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles.* Molecular cell, 2009. **34**(1): p. 47-57.
52. Farnebo, M., *Wrap53, a novel regulator of p53.* Cell cycle, 2009. **8**(15): p. 2343-6.
53. Mahmoudi, S., et al., *Supplemental Data. Wrap53, a Natural p53 Antisense Transcript Required for p53 Induction upon DNA Damage.* Molecular Cell, 2009. **33**(4).
54. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome.* Science, 2005. **309**(5740): p. 1564-6.

55. Munroe, S.H. and J. Zhu, *Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective.* Cellular and molecular life sciences : CMLS, 2006. **63**(18): p. 2102-18.
56. Munroe, S.H., *Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns.* Journal of cellular biochemistry, 2004. **93**(4): p. 664-71.
57. Universal Protein Resource. *Telomerase Cajal body protein 1.* 25.01.2012 28.01.2012]; Available from: http://www.uniprot.org/uniprot/Q9BUR4.
58. Venteicher, A.S. and S.E. Artandi, *TCAB1: driving telomerase to Cajal bodies.* Cell cycle, 2009. **8**(9): p. 1329-31.
59. Cristofari, G., et al., *Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation.* Molecular cell, 2007. **27**(6): p. 882-9.
60. Zhang, Q., N.K. Kim, and J. Feigon, *Architecture of human telomerase RNA.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(51): p. 20325-32.
61. Cioce, M. and A.I. Lamond, *Cajal bodies: a long history of discovery.* Annual review of cell and developmental biology, 2005. **21**: p. 105-31.
62. Xu, C. and J. Min, *Structure and function of WD40 domain proteins.* Protein & cell, 2011. **2**(3): p. 202-14.
63. Mahmoudi, S., et al., *WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies.* PLoS biology, 2010. **8**(11): p. e1000521.
64. Zhong, F., et al., *Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita.* Genes & development, 2011. **25**(1): p. 11-6.
65. Garcia-Closas, M., et al., *Common genetic variation in TP53 and its flanking genes, WDR79 and ATP1B2, and susceptibility to breast cancer.* International journal of cancer, 2007. **121**(11): p. 2532-8.
66. Schildkraut, J.M., et al., *Single nucleotide polymorphisms in the TP53 region and susceptibility to invasive epithelial ovarian cancer.* Cancer research, 2009. **69**(6): p. 2349-57.
67. 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, 2009. **4**(7): p. e6146.
68. Burdall, S.E., et al., *Breast cancer cell lines: friend or foe?* Breast cancer research : BCR, 2003. **5**(2): p. 89-95.
69. ATCC®. *Product Information Sheet for ATCC® HTB-22™.* 2009 2009 14.02.2012]; Available from: http://www.lgcstandards-atcc.org/attachments/17392.pdf.
70. Cailleau, R., M. Olive, and Q.V. Cruciger, *Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization.* In vitro, 1978. **14**(11): p. 911-5.
71. ATCC®. *Product Information Sheet for ATCC HTB-26™.* 2007 2007 14.03.2012]; Available from: http://www.lgcstandards-atcc.org/attachments/17400.pdf.
72. IARC, *IARC TP53 Database, Cell line search*, 2012.
73. Sjøberg, N.O., *Molekylær genetikk*2006: Forlaget Vett & Viten AS.
74. Applied Biosystems, *Model 340A Nucleic Acid Extractor Users Manual*, 1987.
75. Thermo Fisher Scientific Inc, *NanoDrop 1000 Spectrophotometer V3.7 User's Manual*, 2008.
76. Froeyland, C.J. and E.U. Due, *Resequencing of the TP53 gene on ABI 3730 DNA Analyzer*, 2008.
77. Applied Biosystems. *BigDye® Direct Cycle Sequencing Kit Protocol*. 2011 26.08.2011]; Available from: http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_091370.pdf.
78. Mahmoudi, S., *WRAP53 unwrapped; roles in nuclear architechture and cancer*, in *Department of Oncology and Pathology, Cancer Center Karolinska, Karolinska Institutet* 2011: Stockholm.
79. Applied Biosystems, *Applied Biosystems 3730/3730xl DNA Analyzer; Getting Startet Guide*2005.
80. Biotium. *Product Information: GelRed™ Nucleic Acid Gel Stain*. 2011 21.12.2011 31.03.2012]; Available from: http://www.biotium.com/product/product_info/Protocol/PI-41003.pdf.
81. Applied Biosystems. *Applied Biosystems BigDye® XTerminator™ Purification Kit Protocol* 2007 07.2007 06.10.2011]; Available from: http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042772.pdf.

82. Applied Biosystems. *DNA Sequencing by Capillary Electrophoresis. Applied Biosystems Chemistry Guide | Second Edition*. 2009 01.04.2012]; Available from: http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_041003.pdf.
83. Walker, J.M. and R. Rapley, *Molecular Biomethods Handbook*, 2008, Humana Press.
84. Applied Biosystems. *Frequently Asked Questions SeqScape® Software Version 2.7*. 2009 04.2009 01.04.2012]; Available from: http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_064540.pdf.
85. National Center for Biotechnology Information. *Homo sapiens WD repeat containing, antisense to TP53 (WRAP53), RefSeqGene on chromosome 17*. 15.03.2012]; Available from: http://www.ncbi.nlm.nih.gov/nuccore/NG_028245.1?from=5001&to=22432&report=genbank .
86. Pallant, J., *SPSS Survival Manual*. 3 ed2007: The McGraw-Hill Companies.
87. Broad Insitute. *Haploview*. 2010-2012 10.05.2012]; Available from: http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview.
88. National Cancer Institute. *HapMap Linkage Disequilibrium - Phase II - from phased genotypes* 18.05.2012]; Available from: https://cgwb.nci.nih.gov/cgi-bin/hgTrackUi?hgsid=97559&c=chr7&g=hapmapLdPh.
89. Nature Education. *Hardy-Weinberg equilibrium*. 2011 08.05.2012]; Available from: http://www.nature.com/scitable/definition/hardy-weinberg-equilibrium-122.
90. National Cancer Institute. *SNP500Cancer*. 12.03.2012]; Available from: http://variantgps.nci.nih.gov/cgfseq/pages/home.do.
91. National Center for Biotechnology Information. *dbSNP Short Genetic Variations*. 12.03.2012]; Available from: http://www.ncbi.nlm.nih.gov/snp/.
92. Hedström, E., *Cell culturing and siRNA transfection*, 2012.
93. Qiagen. *HiPerFect Transfection Reagent*. 2003-2012 03.04.2012]; Available from: http://www.qiagen.com/products/transfection/transfectionreagents/hiperfecttransfectionreagent.aspx.
94. Wasungu, L. and D. Hoekstra, *Cationic lipids, lipoplexes and intracellular delivery of genes.* Journal of controlled release : official journal of the Controlled Release Society, 2006. **116**(2): p. 255-64.
95. Simms, D., P.E. Cizdziel, and P. Chomczynski, *TRIZOL™: a new reagent for optimal single-step isolation of RNA* in *Focus*®1993, Life Technologies, Inc.
96. Life Technologies Corporation. *TRIzol® Reagent*. 2010 03.04.2012]; Available from: http://tools.invitrogen.com/content/sfs/manuals/trizol_reagent.pdf.
97. Agilent Technologies. *Agilent SurePrint Technology, Content centered microarray design enabling speed and flexibility*. 2003 04.04.2012]; Available from: http://www.chem.agilent.com/Library/technicaloverviews/Public/5988-8171en.pdf.
98. Integrated DNA Technologies *Chemical Synthesis and Purification of Oligonucleotides*. 2011.
99. Agilent Technologies. *Agilent One Color RNA Spike-In Kit*. 2005 12.2005 05.04.2012]; Available from: http://www.chem.agilent.com/Library/usermanuals/Public/5188-5977.pdf.
100. Agilent Technologies. *One-Color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling Protocol Version 6.5, May 2010*. 2010 23.01.2012]; Available from: http://www.chem.agilent.com/Library/usermanuals/Public/G4140-90040_GeneExpression_One-color_v6.5.pdf.
101. Qiagen, *RNeasy® Mini Handbook*, 2010.
102. Agilent Technologies. *Hi-RPM Gene Expression Hybridization Kit*. 2006 16.11.2006 06.04.2012]; Available from: http://www.chem.agilent.com/Library/datasheets/Public/5973-1503_06.11.pdf.
103. Agilent Technologies. *Agilent's DNA Microarray Scanner With SureScan High-Resolution Technology*. 2008 20.05.2008 06.04.2012]; Available from: http://www.chem.agilent.com/Library/brochures/5989-8555en_lo.pdf.
104. Agilent Technologies, *GeneSpring for Gene Expression Analysis, Introduction to GeneSpring GX11*, 2011.
105. Agilent Technologies. *Tutorial: Analyzing Agilent Gene Expression data in GeneSpring GX9*. 08.04.2012]; Available from: http://www.chem.agilent.com/cag/bsp/products/gsgx/downloads/tutorials/genespring_gx9_analyzing_agilent_data_tutorial.pdf.

106. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(9): p. 5116-21.
107. Pawitan, Y., et al., *False discovery rate, sensitivity and sample size for microarray studies.* Bioinformatics, 2005. **21**(13): p. 3017-24.
108. Ingenuity® Systems. *IPA®*. 2012 08.04.2012]; Available from: http://www.ingenuity.com/products/pathways_analysis.html.
109. Ingenuity® Systems. *Calculating and Interpreting the P-values for Functions, Pathways, and Lists in Ingenuity Pathways Analysis*. 2008 03.2008 08.04.2012]; Available from: http://193.191.128.3/images/stories/IPA_PDF/p-values_in_ipa.pdf.
110. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Statistical Society, 1995.
111. Ingenuity® Systems. *IPA Software: Analysis Results Tutorial*.
112. Stanford University. *SOURCE*. 14.03.2012]; Available from: http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch.
113. Norsk Bryst Cancer Gruppe. *Blåboka*. 19.04.2012]; Available from: http://www.nbcg.no/nbcg.blaaboka.html#Anchor-35882.
114. Ingenuity® Systems. *Ingenuity Systems Citation Guidelines*. 2011 01.05.2012]; Available from: http://www.ingenuity.com/library/pdf/ingenuity_citation_guidelines_for_publications.pdf.
115. Oliveros, J.C. *VENNY. An interactive tool for comparing lists with Venn Diagrams*. 2007 28.04.2012]; Available from: http://bioinfogp.cnb.csic.es/tools/venny/index.html.
116. Løvås, G., G *Statistikk for universiteter og høgskoler*. 2 ed2004: Universitetsforlaget.
117. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nature methods, 2008. **5**(1): p. 16-8.
118. Stephens, P.J., et al., *Landscape of cancer genes and mutational processes in breast cancer.* Nature, 2012.
119. Roche. *SNPs: the great importance of small differences*. 16.05.2012]; Available from: http://www.roche.com/pages/facets/22/snps_e.pdf.
120. Hu, J. and P.C. Ng, *Predicting the effects of frameshifting indels.* Genome biology, 2012. **13**(2): p. R9.
121. Tommiska, J., et al., *Breast cancer patients with p53 Pro72 homozygous genotype have a poorer survival.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2005. **11**(14): p. 5098-103.
122. Hoskins, J.M., L.A. Carey, and H.L. McLeod, *CYP2D6 and tamoxifen: DNA matters in breast cancer.* Nature reviews. Cancer, 2009. **9**(8): p. 576-86.
123. Farnebo, M., *WRAP53 siRNA transfection*, 2011.
124. Illumina Inc. *Technical Note: Gene Expression Profiling, The Power of Replicates*. 2012 15.05.2012]; Available from: http://www.illumina.com/documents/products/technotes/technote_power_replicates.pdf.
125. Fundel, K., et al., *Normalization strategies for mRNA expression data in cartilage research.* Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society, 2008. **16**(8): p. 947-55.
126. Fletcher, D., *Percentile normalization*, 2012.
127. Jeanmougin, M., et al., *Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies.* PloS one, 2010. **5**(9): p. e12336.
128. Abcam®. *The Abcam loading control guide*. 1998-2012 03.05.2012]; Available from: http://www.abcam.com/index.html?pageconfig=resource&rid=275.
129. Thermo Scientific. *T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers 260/280 and 260/230 Ratios*. 2009 02.2009 29.04.2012]; Available from: http://www.phenogenomics.ca/transgenics/docs/NanoDrop%20Nucleic-Acid-Purity-Ratios.pdf.
130. Deeb, K.K., D.L. Trump, and C.S. Johnson, *Vitamin D signalling pathways in cancer: potential for anticancer therapeutics.* Nature reviews. Cancer, 2007. **7**(9): p. 684-700.
131. Slattery, M.L., *Vitamin D receptor gene (VDR) associations with cancer.* Nutrition reviews, 2007. **65**(8 Pt 2): p. S102-4.
132. Moll, U.M. and N. Slade, *p63 and p73: roles in development and tumor formation.* Molecular cancer research : MCR, 2004. **2**(7): p. 371-86.

133. National Center for Biotechnology Information. *ATF3 activating transcription factor 3 [Homo sapiens]* 2012 27.04.2012 06.05.2012]; Available from: http://www.ncbi.nlm.nih.gov/gene/467.
134. National Center for Biotechnology Information. *FOSL2 FOS-like antigen 2 [Homo sapiens]* 2012 04.05.2012 06.05.2012]; Available from: http://www.ncbi.nlm.nih.gov/gene/2355.
135. Kawaguchi, M., et al., *IL-17 cytokine family.* The Journal of allergy and clinical immunology, 2004. **114**(6): p. 1265-73; quiz 1274.
136. Zhu, X., et al., *IL-17 expression by breast-cancer-associated macrophages: IL-17 promotes invasiveness of breast cancer cell lines.* Breast cancer research : BCR, 2008. **10**(6): p. R95.
137. Wang, L., et al., *Association analysis of IL-17A and IL-17F polymorphisms in Chinese Han women with breast cancer.* PloS one, 2012. **7**(3): p. e34400.
138. National Center for Biotechnology Information. *ESR1 estrogen receptor 1 [Homo sapiens]* 2012 04.05.2012 06.05.2012]; Available from: http://www.ncbi.nlm.nih.gov/gene/2099.
139. Sledz, C.A., et al., *Activation of the interferon system by short-interfering RNAs.* Nature cell biology, 2003. **5**(9): p. 834-9.
140. Marques, J.T. and B.R. Williams, *Activation of the mammalian immune system by siRNAs.* Nature biotechnology, 2005. **23**(11): p. 1399-405.
141. Leivonen, S.K., *Non-specific effects by siRNA transfections*, 2012.
142. PhosphoSitePlus®. *NR1D1 (human)*. 2003-2012 06.05.2012; Available from: http://www.phosphosite.org/proteinAction.do?id=6044.
143. National Center for Biotechnology Information. *NR3C1 nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor) [Homo sapiens]* 2012 04.05.2012 06.05.2012]; Available from: http://www.ncbi.nlm.nih.gov/gene/2908.

APPENDIX

APPENDIX A: Protocol – NanoDrop® ND-1000 Spectrophotometer

To perform NanoDrop® ND-1000 Spectrophotometer absorbance measurements, the instrument has to be connected to a computer. Turn on the computer and open the instrument software (ND-1000 Software v7.3.1, Thermo Scientific). Clean the optical surfaces using lens-cleaning tissues (Special lens-leaning tissue, Assistent®, No 1019) and initialize the instrument by pipetting 1,0 µl DNase/RNase free water (GIBCO, Ref 10977-035) onto the lower measurement pedestal, lower the sampling arm and press OK using the operating software. Make sure to use the instrument DNA-application (DNA-50) by DNA measurements. Initiate a blank measurement using the DNA dilution agent, and by pressing BLANK. Spectral sample measurements are performed the same way by pipetting 1,0 µl sample, and pressing MEASURE. Keep the samples on ice while measuring and wipe the optical surfaces using lens-cleaning tissue between each measurement to prevent sample carryover. Reblank the instrument for every tenth sample measured. Clean the optical surfaces and print the result report when finished (Thermo Fisher Scientific Inc, 2008).

APPENDIX B: Protocol – BigDye® Direct Cycle Sequencing Kit

Primary reference: BigDye® Direct Cycle Sequencing Kit Protocol (Life Technologies).

Commercial kits were used in the *WRAP53* mutation analysis;

- BigDye® Direct Cycle Sequencing Kit (Applied Biosystems, Part. No. 4458688).
- BigDye® XTerminator™ Purification Kit (Applied Biosystems, Part. No. 4376484).

The BigDye® Direct Cycle Sequencing Kit and 3730 DNA Analyzer operation procedures involve contact with possible hazardous compounds. The XTerminator™ Solution and 10x 3730 Buffer with EDTA are irritant to eyes, respiratory system and skin, and the EDTA buffer might as well cause eye damage. The SAM™ Solution and POP-7™ Performance Optimized Polymer are irritants to eyes and skin. Chemicals should be handled with care and by wearing appropriate personal safety equipment. Use gloves all times to protect the samples.

STEP 1: PCR amplification

1.1) For each forward or reverse reaction, add the components as displayed in table 1 to an appropriate reaction plate (MicroAmp™ Optical 96-Well Reaction Plate with Barcode, Applied Biosystems). Use 1,0 µl DNase/RNase free H_2O instead of genomic DNA to include negative controls.

Table 1: Component concentrations and volumes required to PCR amplify one sample

Components	Volume
Genomic DNA (5 ng/µl)	1,0 µl
WRAP53 M13-tailed Fwd PCR primer (0,8 µM)	0,75µl
WRAP53 M13-tailed Rev PCR primer (0,8 µM)	0,75 µl
BigDye® Direct PCR Master Mix	5,0 µl
DNase/RNase free H_2O (GIBCO, Ref 10977-035)	2,5 µl
Total volume for each reaction	**10,0 µl**

1.2) Mix the components well by pipetting up and down, seal the plate with caps (Domed Cap Strip, Thermo Scientific) and centrifuge the reaction plate briefly.

1.3) Run the PCR amplification in a thermal cycler as displayed in table 2.

Table 2: Time and temperature conditions during PCR amplification

Stage	Tetrad 2 Thermal Cycler (Bio-Rad)	
	Temp	**Time**
Hold	96°C	5 min
Cycle (35 cycles)	94°C	30 sec
	63°C	45 sec
	68°C	45 sec
Hold	72°C	2 min
Hold	4°C	∞

STEP 2: Agarose gel electrophoresis (optional)

2.1) Prepare the agarose gel (recipe in Appendix F). When the gel is hardened, transfer it to the gel chamber and cover it with 1 x TAE buffer (recipe in Appendix F).

2.2) Mix 2,0 µl PCR-product and 2,0 µl 0,1 % bromophenol blue gel loading buffer (recipe in Appendix F) and load the whole volume into a gel well.

2.3) Load 2,0 µl DNA ladder (φX 174-Hae III digest, TaKaRa) into the first well in each row of sample loaded gel wells.

2.4) Run the electrophoresis at 200 V for 30 min, and visualize the results by UV irradiation (GeneGenius Bio Imaging System and GeneSnap Software v7.01.07, Syngene).

Stopping point 1: PCR products can be stored at 4°C over night or at –15°C or –25°C for long-term storage.

STEP 3: Cycle sequencing

3.1) For each forward or reverse reaction, mix the components as displayed in table 3 in an appropriate tube. Mix well and centrifuge the tubes briefly. Keep the premix on ice and in the dark using aluminum foil to avoid fluorescence bleaching.

Table 3: Sequencing reaction mix components and volumes required to cycle sequence one sample

Components	Volume for each reaction
BigDye® Direct Sequencing Master Mix	2,0 µl
One Sequencing primer: • BigDye® Direct M13 Fwd Primer **or** • BigDye® Direct M13 Rev Primer	1,0 µl
Total volume for each reaction	**3,0 µl**

3.2) Add 3,0 µl sequencing reaction mix to each well in the respective forward or reverse reaction plate. Keep reaction plates in the dark.

3.3) Seal the plate with caps and centrifuge the plate briefly.

3.4) Run the sequencing reactions in a thermal cycler as displayed in table 4.

Table 4: Time and temperature conditions during cycle sequencing

Stage	Tetrad 2 Thermal Cycler (Bio-Rad)	
	Temp	**Time**
Hold	37°C	15 min
Hold	80°C	2 min
Hold	96°C	1 min
Cycle (25 cycles)	96°C	10 sec
	50°C	5 sec
	60°C	4 min
Hold	4°C	∞

Stopping point 2: The reaction plate can be stored at 4°C over night, or at –15°C or –25°C for long-term storage.

STEP 4: Purify the sequencing products

Remember to keep the reaction plates in the dark.

4.1) Centrifuge the reaction plate at 100 x G for 1 min.

4.2) Premix the SAM™ Solution and XTerminator™ Solution as displayed in table 5 in an appropriate tube.

Before using the BigDye® XTerminator™ Purification Kit:

- Make sure there are no particles in the SAM™ Solution. Heat the solution to 37°C and resuspend the solution if particles are present.
- Homogenize the XTerminator™ Solution using a high-speed vortexer for 10 sec. Avoid pipetting from the top of the liquid due to rapid sedimentation.
- Use a sterile scalpel and cut the pipette tips to create wide-bore tips (orifice > 1,0 mm) to aspirate the XTerminator™ Solution.
- Calculate to use 15% more of the SAM™ and XTerminator™ Solution than needed by the number of samples to be purified through dead volume. The ratio of SAM™ and XTerminator™ Solution should be 4,5:1 (v/v).

Table 5: BigDye® XTerminator™ Purification Kit components and volumes to purify 96 samples

Components	Volume for each well	Volume for 96 wells
SAM™ Solution	45 µl	4968 µl
XTerminator™ Solution	10 µl	1104 µl
Total volume	**55 µl**	**6072 µl**

4.3) Add 55 µl of the SAM™ and XTerminator™ Solution premix to each sample. Mix the solution in between to prevent bead sedimentation. Seal the plate using caps.

4.4) Vortex the reaction plate for 30 min at 2000 rpm (Illumina High-Speed Microplate Shaker).

4.5) Centrifuge the reaction plate at 1000 x G for 2 min.

Stopping point 3: Sealed reaction plates can be stored up to 48 hours in room temperature or up to 10 days at 4°C or –20°C before proceeding with capillary electrophoresis.

For further information and details, see the BigDye® Direct Cycle Sequencing Kit Protocol (Life Technologies).

APPENDIX C: Protocol – Microarray gene expression analysis

Primary reference: One-Color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling Protocol version 6.5, May 2010 (Agilent Technologies).
Microarrays: SurePrint G3 Hmn GE 80x60K Microarray Kit (Agilent Technologies, Cat. No. G4851-60510).

The microarray procedure involves contact with possible hazardous compounds. Cy3 is a possible carcinogen. The 2x Hi-RPM Hybridization Buffer contains lithium chloride and lithium dodecyl sulfate. Lithium chloride is toxic and a potential teratogen, while lithium dodecyl sulfate is harmful by inhalation and irritating to eyes, skin and the respiratory system. Triton is a component in the 2x Hi-RPM Hybridization Buffer and in the Gene Expression Wash Buffer 1 and 2, and is harmful if swallowed and a risk of serious eye damage. All chemicals should be handled with care and by wearing appropriate personal safety equipment. Wear gloves at all times to protect the samples, the delicate microarrays and yourself.

Sample preparation:

RNA sample concentration: 50 ng/µl RNA

STEP 1: Spike-mix preparation

1.1) Set heat blocks to 37°C, 65°C and 80°C, and water bath to 40°C.

1.2) Heat the Spike-mix at 37°C for a few minutes, and centrifuge the tube briefly.

1.3) Prepare second dilution (1:25) Spike-mix in a new tube

Dilution Buffer	48 µl
Spike-mix	2 µl
Vortex and centrifuge the tube briefly	

1.4) Prepare third dilution (1:20) Spike-Mix in a new tube

Dilution Buffer	38 µl
Second dilution	2 µl
Vortex and centrifuge the tube briefly	

STEP 2: Prepare labeling reaction

2.1) Prepare the T7 Promoter Primer Mix in a new tube on ice

Nuclease-free water	5 µl
T7 Promoter Primer Mix	8 µl

Vortex and centrifuge the tube briefly

2.2) Prepare the samples in new tubes on ice

Sample (50 ng/µl RNA)	2 µl
Third dilution Spike-mix	2 µl

Mix the tube gently and centrifuge briefly

T7 Promoter Primer Mix	1,3 µl

Mix the tube gently and centrifuge briefly

Place the sample tubes at 65°C for 10 min

Place the sample tubes on ice for 5 min

Centrifuge the sample tubes briefly

2.3) **Meanwhile:** prewarm the 5x First Strand Buffer at 80°C for 3-4 min, and keep it in room temperature.

2.4) Prepare the cDNA Master Mix in a new tube

5x First Strand Buffer	20 µl
0,1M DTT	10 µl
10 mM dNTP mix	5 µl
AffinityScript RNase Block Mix	12 µl

2.5) Add 4,7 µl of the cDNA Master Mix to each sample tube.
Mix well by pipetting up and down.

2.6) Place the sample tubes in the 40°C water bath for 2 hours.

2.7) Meanwhile: Set heat block to 70°C.

2.8) Place the sample tubes at 70°C for 15 min.

2.9) Place the sample tubes on ice for 5 min.

2.10) Centrifuge the sample tubes briefly.

Stopping point 1: move the sample tubes to –80°C.

2.11) Prepare the Transcription Master Mix in a new tube

5x Transcription Buffer	32 µl
Nuclease-free water	7,5 µl
NTP Mix	10 µl
0,1 M DTT	6 µl
T7 RNA Polymerase Blend	2,1 µl

Cyanine 3-CTP (keep in dark) 2,4 µl

From here: keep the sample tubes dark at all times using aluminum foil!

2.12) Add 6 µl of the Transcription Master Mix to each sample tube.

2.13) Place the sample tubes in the 40°C water bath for 2 hours.

2.14) Centrifuge the sample tubes briefly.

Stopping point 2: move the sample tubes to –80°C.

STEP 3: Purify labeled and amplified cRNA

3.1) Pre-cool the centrifuge to 4°C.

3.2) Purify the samples using the RNeasy® Mini Kit (Qiagen).

Work in room temperature, as quick and dark as possible.

Add the following solution to the sample tubes:

DNase/RNase free water	84 µl
RLT buffer	350 µl
Mix by pipetting	
Ethanol (100%)	250 µl
Mix by pipetting	

3.3) Transfer the whole sample volume (700 µl) to an RNeasy mini column.

3.4) Centrifuge at 4°C and 13000 rpm for 30 sec. Discard the flow-through.

3.5) Add 500 µl RPE Buffer to each spin column.

3.6) Centrifuge at 4°C and 13000 rpm for 30 sec. Discard the flow-through.

3.7) Add 500 µl RPE Buffer to each spin column.

3.8) Centrifuge at 4°C and 13000 rpm for 60 sec. Discard the flow-through.

3.9) Centrifuge at 4°C and 13000 rpm for 30 sec.

3.10) Place the spin columns in new tubes.

3.11) Add 30 µl DNase/RNase-free water to each spin column.

Incubate for 1 min.

Centrifuge at 4°C and 13000 rpm for 30 sec.

STEP 4: cRNA quantification

Quantify the RNA using a NanoDrop® ND-1000 Spectrophotometer (Saveen Werner). Use the same DNase/RNase-free water as the cRNA is eluted in to blank the spectrophotometer. The DNase/RNase-free water should also be measured. Remember to keep the samples dark and on ice while measuring. Before proceeding to the hybridization step (step 5), the cRNA yield and specific activity results should be examined. A cRNA yield of > 1,65 µg and specific activity of > 9,0 pmol Cy3 per µg cRNA are required in the hybridization step. Repeat the cRNA preparation if the results deviates the requirements.

$$\textbf{µg of cRNA} = \frac{\text{(Concentration of cRNA (ng/µl)} \times \text{30 µl (elution volume)}}{1000}$$

$$\textbf{pmol Cy3 per µg cRNA} = \frac{\text{Concentration of Cy3 (pmol/µl)}}{\text{Concentration of cRNA (ng/µl)}} \times 1000$$

In the hybridization step, 600 ng cRNA is added to the reaction. Use the spectrophotometric measurements to calculate the required sample volume;
Sample volume (µl) = 600 ng / cRNA concentration (ng/µl)

According to the procedure, the cRNA sample volume and volume DNase/RNase free water added should not exceed 19 µl. Adjust the water volume in accordance to the sample volume;
Volume water (µl) = 19,0 µl – cRNA sample volume (µl)

Stopping point 3: move the sample tubes to –80°C.

STEP 5: Hybridization

Turn on the hybridization oven to 65°C at least one hour before use.

5.1) Set heat block to 60°C.

5.2) Prewarm the 10x Blocking Agent at 37°C for 3-4 min.

5.3) Mix the following components in new sample tubes

cRNA	600 ng (calculate the required volume in µl)
10x Blocking Agent	5 µl
DNase/RNase-free water	x µl (adjust to added sample volume)
Total	**24 µl**
Fragmentation Buffer	1 µl
Total sample volume	**25 µl**

5.4) Place the sample tubes at 60°C for 30 min.

5.5) Place the sample tubes on ice for 1 min.

5.6) Add 25 µl 2x Hi-RPM Hybridization Buffer to each sample tube.
Mix by careful pipetting.
Centrifuge in room temperature at 13000 rpm for 1 min. Make sure there is no air bubbles percent. Centrifuge again if necessary to eliminate bubbles.

5.7) Place the samples on ice, and apply the samples to the array.
Use the hybridization gasket slides (Agilent Technologies) to apply the samples. Further apply the SurePrint G3 Hmn GE 80x60K Microarray (Agilent Technologies) with the probes in contact with the samples. Make sure there are no stationary air bubbles impairing hybridized array quality.

5.8) Place the array in the hybridization oven at 65°C and 10 rpm for 17 hours.

STEP 6: Washing and scanning the array

It is highly recommended to add 0,005% Triton X-102 (Agilent Technologies) to the Gene Expression Wash Buffer 1 and 2 (Agilent Technologies) prior to use to reduce the possibility of array wash artifacts. Keep the buffer trays on magnet stirrers to keep fluids in motion for improved array quality.

6.1) Prewarm Buffer 2 to 37°C.

6.2) Wash the array for 1 min in room tempered Buffer 1.

6.3) Wash the array for 1 min in the 37°C Buffer 2.

6.4) Slowly lift the array up from Buffer 2 to prevent the formation of buffer droplets on the array.

6.5) Place the array in a dark, air-protected box and scan as soon as possible.

For further information and details, see the One-Color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling Protocol version 6.5, May 2010 (Agilent Technologies).

APPENDIX D: Results from the *WRAP53* mutation analysis

Table 10: *WRAP53* sequence alterations detected in the ULL tumor and blood samples. Sequenced ULL blood samples are marked *. BC = base change, 1 = heterozygous genotype, 2 = minor homozygous genotype, - = data not available.

Sample ID	Location	Codon	Base change	Coding description	Codon change	Amino acid (predicted)	Protein change	Type	Geno-type
ULL-T-003	Intron 5	-	C>T	c.731+27C>T	-	-	-	BC	1
ULL-T-007	Exon 2*	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-009	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-010	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-011	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-014	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-017	Exon 1β*	-	G>C	c.–245G>C	-	-	-	BC	1
ULL-T-020	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-023	Intron 5	-	C>T	c.731+27C>T	-	-	-	BC	1
ULL-T-024	Intron 6	-	C>T	c.823–10C>T	-	-	-	BC	1
ULL-T-025	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-026	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-027	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-028	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11*	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-036	Exon 2*	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-037	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3*	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-038	Exon 1β*	-	G>C	c.–245G>C	-	-	-	BC	1
ULL-T-044	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-053	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-055	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-063	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-064	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-065	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-066	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-069	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-071	Intron 6	-	C>T	c.823–10C>T	-	-	-	BC	1
ULL-T-072	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1

ULL-T-074	Exon 11*	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-080	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-082	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-083	Exon 1β	-	G>C	c.–245G>C	-	-	-	BC	1
ULL-T-084	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3*	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-085	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-088	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-090	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-093	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-094	Intron 6	-	C>T	c.823–10C>T	-	-	-	BC	1
ULL-T-097	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-098	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-100	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3*	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-106	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-107	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-109	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
ULL-T-111	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-113	Intron 2	-	C>G	c.432–15C>G	-	-	-	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-115	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-134	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-137	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-138	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-139	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-141	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-142	Exon 11	522	-	c.1566_1567insG or c.1564_1567delCGinsGGG				Indel	-
ULL-T-143	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-152	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-155	Exon 2*	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-162	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-163	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	2
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	2
	Intron 8	-	G>A	c.1165–30G>A	-	-	-	BC	1
	Exon 10	436	T>C	c.1308 T>C	GCT>GCC	Ala>Ala	p.A436A	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-165	Intron 3	-	G>A	c.530+17G>A	-	-	-	BC	1
ULL-T-171	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-176	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-180	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-184	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1

ULL-T-186	Exon 1β	-	G>C	c.–245G>C	-	-	-	BC	1
ULL-T-196	Intron 5	-	C>T	c.731+27C>T	-	-	-	BC	1
ULL-T-198	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-199	Exon 11*	522	G>A	c.1566 G>A	GCG>GCA	Ala>Ala	p.A522A	BC	1
ULL-T-201	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Intron 2	-	C>G	c.432–15C>G	-	-	-	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-202	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-209	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
	Intron 6	-	C>G	c.823–10C>T	-	-	-	BC	1
ULL-T-211	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-212	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-218	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-225	Exon 11*	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-227	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-236	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-237	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-242	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-247	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-253	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11*	522	-	c.1565_1568delGC			-	Indel	1
ULL-T-256	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-262	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-263	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 3	150	C>T	c.450 C>T	TTC>TTT	Phe>Phe	p.F150F	BC	1
	Exon 11*	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-268	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	2
ULL-T-270	Exon 1β	-	G>C	c.–245G>C	-	-	-	BC	1
ULL-T-277	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11*	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1
ULL-T-279	Exon 2	68	C>G	c.202 C>G	CGG>GGG	Arg>Gly	p.R68G	BC	1
	Exon 11	522	C>G	c.1565 C>G	GCG>GGG	Ala>Gly	p.A522G	BC	1

APPENDIX E: Gene lists from the gene expression study

Table 11: Significant upregulated genes in the MCF-7 cell line 72 hours after siWRAP53#2 transfection

ID	Gene Name	Fold Change	ID	Gene Name	Fold Change	ID	Gene Name	Fold Change
1	MMP1	25,8	51	GJB4	4,3	101	THBD	3,2
2	RGS9	16,0	52	CYP2C8	4,3	102	ANTXR1	3,2
3	SPOCK1	11,3	53	EPHB6	4,3	103	KLK6	3,2
4	ITGB8	11,1	54	TNFRSF21	4,2	104	MTL5	3,2
5	FOXQ1	10,9	55	MGC16075	4,2	105	PSG1	3,2
6	SPRR1A	9,4	56	SPINK1	4,2	106	TGM2	3,1
7	ATF3	9,0	57	HES2	4,1	107	ZNF608	3,1
8	FLJ22536	7,5	58	NUPR1	4,1	108	PSG8	3,1
9	KLHDC7B	7,5	59	ACTA2	4,0	109	GRB10	3,1
10	KALRN	7,4	60	CNN3	4,0	110	SLC14A1	3,1
11	NCF2	7,3	61	CDKN2B	3,9	111	WSCD1	3,1
12	PXDC1	7,3	62	PSG2	3,9	112	PTGER3	3,1
13	GPR87	7,3	63	SERPINB5	3,9	113	AQP3	3,1
14	NCRNA00324	7,2	64	IFIT2	3,8	114	DUSP10	3,1
15	ALOX5	7,2	65	S100A2	3,8	115	LOXL4	3,1
16	TCHHL1	7,1	66	KLHL24	3,8	116	NEDD9	3,1
17	MALL	7,1	67	PTPRH	3,8	117	SEMA7A	3,1
18	NTN4	7,0	68	PHLDA1	3,8	118	PHF21B	3,0
19	ARL14	6,5	69	CCL26	3,8	119	FYN	3,0
20	CLDN1	6,5	70	KLK7	3,7	120	AKR1B15	3,0
21	INHBA	6,4	71	GABARAPL1	3,7	121	CLSTN2	3,0
22	MMP24	6,3	72	PID1	3,7	122	CENPV	3,0
23	GEM	6,2	73	SERPINE2	3,6	123	CLIP2	3,0
24	RASD1	6,0	74	RHCG	3,6	124	COL5A1	3,0
25	FLJ13197	5,9	75	TNFSF15	3,6	125	NOG	3,0
26	IVL	5,9	76	DSCAML1	3,6	126	RASGRP1	3,0
27	FHL2	5,9	77	DDX60L	3,6	127	CAPN13	3,0
28	FBXO32	5,8	78	BCAT1	3,6	128	SPRR3	3,0
29	F2R	5,6	79	SHISA2	3,6	129	KIAA0226L	2,9
30	SPRR1B	5,6	80	FAS	3,6	130	RFTN2	2,9
31	NEURL3	5,6	81	FSIP2	3,5	131	PSG9	2,9
32	CACNG6	5,5	82	WDR66	3,5	132	PLAC1	2,9
33	LYPD1	5,4	83	ENTPD3	3,5	133	COL20A1	2,9
34	OLFML3	5,4	84	RCAN2	3,5	134	MAP2	2,9
35	MMP13	5,3	85	OXTR	3,5	135	TGFBI	2,9
36	HTR1F	5,3	86	AKR1B10	3,5	136	PMAIP1	2,9
37	DHRS3	5,2	87	SLIT2	3,5	137	FAM155A	2,9
38	WLS	5,2	88	GADD45A	3,5	138	SIGLEC15	2,9
39	NHS	5,2	89	ENOX1	3,4	139	PALLD	2,9

40	DNM3	4,9	90	LGR6	3,4	140	PSORS1C1	2,8
41	PLXNA2	4,9	91	FBN2	3,4	141	RGS20	2,8
42	KIAA1239	4,8	92	PMEPA1	3,4	142	FAM125B	2,8
43	LAMC2	4,7	93	MEF2C	3,4	143	ROR1	2,8
44	MGC20647	4,6	94	ANKRD29	3,4	144	IRAK2	2,8
45	LRRC4	4,5	95	STAT4	3,3	145	RAB7B	2,8
46	SAMD4A	4,5	96	ACOXL	3,3	146	CST6	2,8
47	CHAC1	4,4	97	CCDC80	3,3	147	DDIT3	2,8
48	COL17A1	4,4	98	LTB	3,3	148	MGLL	2,8
49	PGM5	4,4	99	LAMB3	3,3	149	SLC7A11	2,8
50	KIAA1683	4,3	100	ZNF532	3,3	150	CTNNA3	2,8

ID	Gene Name	Fold Change	ID	Gene Name	Fold Change	ID	Gene Name	Fold Change
151	APOBEC3H	2,8	201	VWCE	2,4	251	AP1S2	2,2
152	BBC3	2,8	202	CDH11	2,4	252	RHOU	2,1
153	TLR2	2,7	203	ARHGAP36	2,4	253	PLD1	2,1
154	UPP1	2,7	204	CYP1B1	2,4	254	PGM2L1	2,1
155	CLIP4	2,7	205	VEPH1	2,4	255	OSBPL5	2,1
156	CAPN8	2,7	206	TP53TG1	2,4	256	LHPP	2,1
157	TGFB2	2,7	207	LAMA3	2,4	257	ARHGDIB	2,1
158	PDLIM3	2,7	208	CTF1	2,4	258	LCAT	2,1
159	ABCA1	2,7	209	INPP1	2,4	259	RND3	2,1
160	SPATA18	2,7	210	DIO3OS	2,4	260	RELB	2,1
161	HCN4	2,7	211	FLJ40852	2,4	261	AHR	2,1
162	GPNMB	2,6	212	LONRF3	2,4	262	AES	2,1
163	F2RL1	2,6	213	SLC6A8	2,4	263	STAU2	2,1
164	DAPK2	2,6	214	GSTA4	2,3	264	HRK	2,1
165	MBP	2,6	215	PAPLN	2,3	265	NDRG4	2,1
166	FOSL1	2,6	216	CCK	2,3	266	S100A6	2,1
167	MAF	2,6	217	MME	2,3	267	SLC17A5	2,1
168	IL6	2,6	218	CMAHP	2,3	268	PROC	2,0
169	TM4SF1	2,6	219	WNT4	2,3	269	BTG1	2,0
170	SLC5A10	2,6	220	FRAS1	2,3	270	GLP2R	2,0
171	TXLNB	2,6	221	ZFP36L1	2,3	271	TNFRSF11B	2,0
172	ANK1	2,6	222	ABLIM2	2,3	272	ITGA6	2,0
173	STAMBPL1	2,6	223	ADM2	2,3	273	SEMA5A	2,0
174	TIMP3	2,6	224	ASNS	2,3	274	HS3ST1	2,0
175	FAM196A	2,6	225	LHFPL2	2,3	275	ACSS1	2,0
176	PSG10P	2,5	226	NXNL2	2,3	276	BEX2	2,0
177	KSR1	2,5	227	FSTL3	2,3	277	XKR6	2,0
178	RBP1	2,5	228	RUNX2	2,3	278	GLS2	2,0
179	CACNG1	2,5	229	SLC16A14	2,3	279	IQCJ-SCHIP1	2,0

180	ZNF469	2,5	230	CLCF1	2,3	280	WDFY4	2,0
181	SLCO2A1	2,5	231	BEND7	2,3	281	EYA2	2,0
182	XYLT1	2,5	232	JAKMIP2	2,2	282	RNU1-5	2,0
183	CAPS2	2,5	233	TFPI	2,2	283	S100A9	2,0
184	HSPG2	2,5	234	UCHL1	2,2	284	PRICKLE2	2,0
185	GULP1	2,5	235	HAPLN3	2,2	285	FAM71E1	2,0
186	CCDC146	2,5	236	CDH26	2,2	286	ProSAPiP1	2,0
187	TRIM29	2,5	237	MFGE8	2,2	287	PCBP4	2,0
188	ALDH1A3	2,5	238	FAM189A2	2,2	288	EDA2R	2,0
189	TMEM45B	2,5	239	CYB5RL	2,2	289	SLC25A21	2,0
190	SLC13A3	2,5	240	EVPLL	2,2	290	SLC12A4	2,0
191	ANXA1	2,5	241	SESN3	2,2	291	EPDR1	1,9
192	SOX4	2,5	242	TNFSF9	2,2	292	TNFRSF10C	1,9
193	LAMP3	2,5	243	FAM107B	2,2	293	RHOQ	1,9
194	PRSS23	2,4	244	ATP1A1OS	2,2	294	AGA	1,9
195	EPAS1	2,4	245	AUTS2	2,2	295	CROT	1,9
196	AKR1B1	2,4	246	TNFAIP3	2,2	296	XG	1,9
197	SLC23A3	2,4	247	POPDC2	2,2	297	FAM105A	1,9
198	KCNK2	2,4	248	CDKN1A	2,2	298	TUFT1	1,9
199	MUCL1	2,4	249	FDXR	2,2	299	ST3GAL1	1,9
200	PLEKHH2	2,4	250	LMO7	2,2			

Table 12: Significant upregulated genes in the MDA-MB-231 cell line 72 hours after siWRAP53#2 transfection

ID	Gene Name	Fold Change	ID	Gene Name	Fold Change	ID	Gene Name	Fold Change
1	VTCN1	5,3	51	POLR2F	2,2	101	SCD5	1,9
2	SELV	4,3	52	AADAC	2,2	102	HOPX	1,9
3	HOXB8	3,8	53	SLC10A2	2,2	103	BMP2	1,9
4	CCDC129	3,6	54	FAM22A	2,2	104	SNRPD3	1,9
5	HNF1A	3,6	55	SLITRK6	2,2	105	GPR64	1,9
6	ZNF847P	3,5	56	FTCD	2,2	106	SLC16A14	1,9
7	EYS	3,4	57	SMCR5	2,1	107	DIP2C	1,9
8	SYCE3	3,2	58	TTLL9	2,1	108	EREG	1,8
9	EDA2R	3,2	59	IL13RA2	2,1	109	RINL	1,8
10	ENKUR	3,1	60	FAM166A	2,1	110	PARM1	1,8
11	PHACTR1	3,1	61	UCHL1	2,1	111	PSG5	1,8
12	PTGS2	3,1	62	PPP1R1C	2,1	112	IGKV1-5	1,8
13	TDO2	3,0	63	ZIM2	2,1	113	MATL2963	1,8
14	LZTS1	3,0	64	TEX12	2,1	114	PANX2	1,8
15	NECAB2	2,9	65	NCRNA00167	2,1	115	PPEF1	1,8
16	LAMA1	2,8	66	FAM83B	2,0	116	KAZALD1	1,8
17	PTX4	2,8	67	FERMT1	2,0	117	LRRC7	1,8
18	WWTR1	2,8	68	MED12L	2,0	118	FGFR1	1,8

19	KIAA0319	2,8	69	FAM71A	2,0	119	SLC25A34	1,8
20	CRCT1	2,8	70	HAS2	2,0	120	UTS2	1,8
21	TEKT3	2,7	71	SSX1	2,0	121	OR10AD1	1,8
22	C21orf116	2,6	72	COL20A1	2,0	122	NPAS3	1,8
23	RGS18	2,6	73	FIGF	2,0	123	IL8	1,8
24	KIAA1841	2,6	74	SLC6A14	2,0	124	ZNF519	1,8
25	NDST3	2,6	75	FAM18A	2,0	125	NRIP3	1,8
26	KLK6	2,6	76	MCF2L	2,0	126	CLEC18C	1,8
27	SPINK14	2,6	77	ARMC2	2,0	127	IL1A	1,8
28	FRG2C	2,6	78	CAPN3	2,0	128	FST	1,8
29	TCP11	2,6	79	SPAG16	2,0	129	SSX3	1,8
30	FLJ12825	2,6	80	TRIM17	2,0	130	TBC1D21	1,8
31	BCOR	2,6	81	SRRM4	2,0	131	FZD8	1,7
32	LRRC17	2,6	82	CXCL1	2,0	132	SNX10	1,7
33	AMZ2P1	2,5	83	EVPLL	2,0	133	ST6GAL1	1,7
34	MARCH1	2,5	84	B4GALNT1	2,0	134	PLD5	1,7
35	DHRS9	2,5	85	SCARNA23	2,0	135	TNFSF9	1,7
36	SNORA46	2,5	86	LTB	2,0	136	PEG10	1,7
37	SPP1	2,4	87	RXFP3	1,9	137	FAM95B1	1,7
38	KCNK16	2,4	88	HPX-2	1,9	138	DUSP13	1,7
39	NPW	2,4	89	MAP2	1,9	139	ZNF204P	1,7
40	ANP32A	2,4	90	SLC13A3	1,9	140	CMAHP	1,7
41	GCNT3	2,3	91	PRO1596	1,9	141	ADAMTS4	1,7
42	KCTD4	2,3	92	CLDN1	1,9	142	SPATA1	1,7
43	DEFB4A	2,3	93	ANKFN1	1,9	143	LAIR2	1,7
44	SERPINB5	2,3	94	CXCL3	1,9	144	ABI3BP	1,7
45	PTPRC	2,3	95	SULF1	1,9	145	ARHGEF9	1,7
46	OLAH	2,3	96	NR5A2	1,9	146	LAD1	1,7
47	TMEM14E	2,3	97	SSX4B	1,9	147	ZNF702P	1,7
48	AKR1C1	2,3	98	CD300LG	1,9	148	DCLK1	1,7
49	CCDC102B	2,3	99	MLLT11	1,9	149	TRPV1	1,7
50	SPRR2G	2,2	100	PCDHB15	1,9	150	PI3	1,7

ID	Gene Name	Fold Change	ID	Gene Name	Fold Change
151	CCL3	1,7	201	MUSK	1,5
152	HS6ST2	1,7	202	SOX4	1,5
153	TNFRSF11B	1,7	203	MPZL2	1,5
154	KIAA0754	1,7	204	MEF2C	1,5
155	ADTRP	1,7	205	ZNF660	1,5
156	ZNF616	1,7	206	DPYSL5	1,5
157	CNTNAP3	1,7	207	PRINS	1,5
158	PELI2	1,7	208	ITGB8	1,5
159	TIE1	1,7	209	SOD2	1,5

160	DNHD1	1,6	210	C1RL	1,5
161	ADAMTS1	1,6	211	FAM27A	1,5
162	STC2	1,6	212	FSIP2	1,5
163	TTC18	1,6	213	COL4A6	1,5
164	MIG7	1,6	214	SULT1C2	1,5
165	KISS1	1,6	215	HSD17B2	1,5
166	RD3	1,6	216	LHPP	1,5
167	LOX	1,6	217	MDFI	1,5
168	CLMP	1,6	218	RGS16	1,5
169	CXCL2	1,6	219	SLC2A14	1,5
170	C1S	1,6	220	COL8A1	1,5
171	CYB5RL	1,6	221	CTSC	1,5
172	CPA3	1,6	222	PIR	1,5
173	MMP1	1,6	223	NPAS1	1,5
174	CDK14	1,6	224	SEMA3A	1,5
175	GTDC1	1,6	225	CD22	1,5
176	AZGP1P1	1,6	226	TCN1	1,5
177	MMEL1	1,6	227	DEFB103B	1,5
178	TM4SF1	1,6	228	BICD2	1,5
179	SEL1L3	1,6	229	ZBTB32	1,5
180	NXF3	1,6	230	CYP1B1	1,5
181	CPA6	1,6	231	FBXO16	1,5
182	HCLS1	1,6	232	PCDHGB4	1,5
183	ATXN1L	1,6	233	SLC2A3	1,5
184	ITGA10	1,6	234	CCDC80	1,5
185	TET3	1,6	235	MESDC2	1,5
186	SSX8	1,6	236	REN	1,5
187	ZMYND8	1,6	237	CACNG6	1,4
188	TMEM144	1,6	238	SLC7A2	1,4
189	MCF2L-AS1	1,6	239	FBXO36	1,4
190	NR3C2	1,5	240	GABARAPL1	1,4
191	NAMPT	1,5	241	ZNF674	1,4
192	SCEL	1,5	242	G3BP2	1,4
193	C1QTNF6	1,5	243	THBD	1,4
194	DDIT4L	1,5	244	CYP2J2	1,4
195	ZNF267	1,5	245	NUPR1	1,4
196	PLCB2	1,5	246	TLCD1	1,4
197	THBS4	1,5	247	RGS2	1,4
198	OGFR	1,5	248	DST	1,4
199	CYBB	1,5	249	NR4A2	1,4
200	OSBPL5	1,5			

APPENDIX F: Reagents

50x TAE buffer

Reagents

Trizma® base, Sigma-Aldrich® Norway AS (Prod. No. T1503)
EDTA disodium salt, BDH® (Prod. 100935V)
Glacial Acetic Acid, MERCK (Cat. No. 100063)

Procedure

Table 13: Component volumes and amounts to make 50x TAE buffer

Components	Amount
Trizma® base	242 g
EDTA disodium salt	100 ml 0,5 M EDTA (pH = 8,0)
Glacial Acetic Acid	57,1 ml
MQ-water	up to 1000 ml

1) Scale in 242 g Trizma® base
2) Add 500 ml MQ-water
3) Add 100 ml 0,5 M EDTA (pH = 8,0) and 57,1 ml Glacial Acetic Acid
4) Add MQ-water to adjust to 1000 ml
Store the buffer in room temperature.

1x TAE buffer

Dilute the 50x TAE buffer 1:10 in MQ-water.
Store the buffer in room temperature.

Agarose gel (1,5% agarose)

Reagents

Agarose, Bio-Rad Laboratories (Cat. No. 161-3102)
GelRed™ Nucleotid Acid Stain, Biotium (Cat. No. 41003-1)
1x TAE buffer

Procedure

Table 14: Component volumes and amounts to make agarose gel (1,5% agarose)

Components	Amount
Agarose	5,25 g
1x TAE buffer	350 ml
GelRed™ NucleicAcid Stain	35,0 µl

1) Mix 5,25 g Agarose and 350 ml 1x TAE buffer in an Erlenmeyer flask
2) Heat to boiling in a microwave
3) Swirl the flask gently until all the agarose is completely dissolved
4) Cool the gel solution to about 65°C
5) Add 35,0 µl GelRed™ Nucleotid Acid Stain (Biotium)
6) Pour the gel and let it harden for 30 min

Store the gel by 4°C. Wrap the gel in plastic foil to avoid drying.

Gel loading buffer (0,1% Bromophenol blue)

Reagents

Bromophenol blue, Bio-Rad Laboratories (Cat. No. 161-0404)
Ficoll® PM 400, Sigma-Aldrich® Norway AS (Prod. No. F4375)
1x TAE buffer

Procedure

Table 15: Component volumes and amounts to make 0,1% Bromophenol blue gel loading buffer

Components	Amount
Bromophenol blue	0,025 g
Ficoll® PM 400	5,0 g
1x TAE buffer	25 ml

1) Scale in 0,025 g Bromophenol blue and 5,0 g Ficoll in a 50 ml tube
2) Add 25 ml 1x TAE buffer
3) Ficoll needs time to dissolve properly, so vortex the tube for about 24 hours to get a homogenous solution

Keep the buffer in room temperature for short-term storage, and by 4°C for long-term storage.

1x Sequencing buffer

Reagents

10x 3730 Buffer with EDTA, Applied Biosystems (Part. No. 4335613)

Procedure

Dilute the 10x 3730 Buffer with EDTA 1:10 in MQ-water.
Store the buffer at 4°C.

APPENDIX G: Chemicals and equipment

Table 16: Chemicals and reagents with supplier and ordering information

Chemical	Supplier	Ordering information
2x Hi-RPM Hybridization Buffer	Agilent Technologies	Part. No. 5188-6420
10x 3730 Buffer with EDTA	Applied Biosystems	Part. No. 4335613
10x GE Blocking Agent	Agilent Technologies	Cat. No. 5188-5973
10% Triton X-102	Agilent Technologies	Part. No. 5188-5903
25x Fragmentation Buffer	Agilent Technologies	Part. No. 5185-5974
φX 174-Hae III digest	TaKaRa	Code No. 3405 A
Absolute alcohol prima	Kemetyl Norge AS	-
Agilent Gene Expression Wash Buffer 1	Agilent Technologies	Part. No. 5188-5325
Agilent Gene Expression Wash Buffer 2	Agilent Technologies	Part. No 5188-5326
Big Dye® Direct Cycle Sequencing Kit	Applied Biosystems	Part. No. 4458688
BigDye® XTerminator™ Purification Kit	Applied Biosystems	Part. No. 4376484
Bradford assay	Bio-Rad Laboratories	Cat. No. 500-0006
Bromophenol blue	Bio-Rad Laboratories	Cat. No. 161-0404
Certified™ Molecular Biology Agarose	Bio-Rad Laboratories	Cat. No. 161-3102
Cyanine 3 CTP Dye Pack	Agilent Technologies	Part. No. 5190-2329
DNase/RNase free H_2O	GIBCO	Ref 10977-35
Dulbecco`s modified Eagle Medium	HyClone, Thermo Scientific	Cat. No. AW-L25546
EDTA disodium salt	BDH®	Prod. 100935V
Fetal bovine serum	HyClone, Thermo Scientific	Cat. No. SV30160.03
Ficoll® PM 400	Sigma-Aldrich® Norway AS	Prod. No. F4375
Glacial Acetic Acid	MERCK	Cat. No. 100063
GelRed™ Nucleic Acid Gel Stain	Biotium	Cat. No. 41003-1
HiPerFect Transfection Reagent	Qiagen	Cat. No. 301704
KAPA SYBR® FAST	Kapa Biosystems	Cat. No. KK4602
LowInput Quick Amp Labeling Kit	Agilent Technologies	Part. No. 5190-2331
Monoclonal Anti-β-actin antibody produced in mouse	Sigma-Aldrich®	Cat. No. A5441
Negative Control siRNA (20 nmol)	Qiagen	Cat. No. 1027310
Plasmocin™	InvivoGen	Cat. code ant-mpp
POP-7™ Performance Optimized Polymer	Applied Biosystems	Part. No. 4363929
RNase Away®	Molecular BioProducts	Cat. No. 7005
RNA Spike-In Kit	Agilent Technologies	Part. No. 5188-5282
RNeasy MinElute CleanUp Kit	Qiagen	Cat. No. 74204
RNeasy® Mini Kit	Qiagen	Cat.No. 74106
Sephadex™ G-50 Superfine	GE Healthcare	Cat. No. 17-0041-01
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Scientific	Cat. No. 34096
Trizma® base	Sigma-Aldrich® Norway AS	Prod. No. T1503
WRAP53 C2 antibody	Innovagen	Cat. No. PA-2020-100
WRAP53 PCR primers	Eurogentec	-
WRAP53 siRNA / Hs_FLJ10385_2	Qiagen	Cat. No. SI003889948

Table 17: Equipment with supplier and ordering information

Equipment	Supplier	Ordering information
96-well plate septa	Applied Biosystems	Cat. No. 4315933
Agilent's DNA Microarray Scanner With SureScan High-Resolution Technology	Agilent Technologies	Model G2565CA
Applied Biosystems 3730 DNA Analyzer	Applied Biosystems	-
DNA Engine Tetrad 2 Peltier Thermal Cycler	Bio-Rad Laboratories	-
Disposable Scalpels (sterile)	Swann-Morton	Ref. 0503
Domed Cap Strip	Thermo Scientific	Cat. No. AB-0602
Eppendorf Centrifuge 5804	Eppendorf	-
Eppendorf Mini Spin®	Eppendorf	-
Eppendorf Centrifuge 5415D	Eppendorf	-
GeneGenius Bio Imaging System	Syngene	-
Hybridization Gasket Slides	Agilent Technologies	Cat. No. G2534-60014
Hybridization oven	SHEL LAB	-
Illumina High-Speed Microplate Shaker	Illumina	-
Lense-cleaning paper, Assistant No.1019	VWR International	Art. No. 763-0319
MicroAmp™ Optical 96-Well Reaction Plate with Barcode	Applied Biosystems	Part. No. 4306737
Multiscreen Column Loader 45 UL	Millipore	Cat. No. CP5SN5099
MultiScreen® HV Filter Plates	Millipore	Cat. No. MAHVN4550
MultiScreen® PCR µ 96	Millipore	Cat. No. LSKMPCR50
NanoDrop® ND-1000 Spectrophotometer	Saveen Werner	-
Powerpac 300 Electrophoresis Power Supply	Bio-Rad Laboratories	-
Safe-Lock Tubes 1,5 ml	Eppendorf	Order No. 0030 123.328
Sub-Cell® Model 192	Bio-Rad Laboratories	-
SurePrint G3 Hmn GE 80x60K Microarray Kit	Agilent Technologies	Cat. No. G4851-60510
Therma-Fast®96, Non-Skirted	Thermo Scientific	Cat. No. AB-0600/G
Software		
Agilent Scan Control	Agilent Technologies	-
GeneSnap Software, Version 7.01.07	Syngene	-
GeneSpring GX 12.0	Agilent Technologies	-
Feature Extraction v10.7.3.1	Agilent Technologies	-
Ingenuity Pathways Analysis	Ingenuity® Systems	-
ND-1000 Software v7.3.1	Thermo Fisher Scientific Inc	-
SeqScape v.2.7	Applied Biosystems	-