Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of Doctor of Natural Sciences

presented by

M.Sc. Julia Bitzer

born in: Leipzig

Oral-examination: 23.06.2015

Molecular characterization of the breast cancer associated antigen NY-BR-1

Referees:

Prof. Dr. Oliver Gruss

Prof. Dr. Dirk Jäger

Danksagung

Als Erstes möchte ich mich bei Herrn Prof. Dr. Dirk Jäger bedanken für die Möglichkeit der Ausführung dieser interessanten Arbeit in seinem Labor.

Mein Dank für die akademische Betreuung und wertvollen Hinweise für die Arbeit gilt meinem Doktorvater Herrn Prof. Dr. Oliver Gruss.

Ein herzliches Dankeschön geht an meine Betreuerin Dr. Inka Zörnig für die sehr hilfsbereite wissenschaftliche Betreuung, die neuen Ideen und die vielen lehrreichen Hinweise und Diskussionen.

Natürlich gibt es ein ganz großes Dankeschön für das Laborteam: Iris Kaiser, Claudia Ziegelmeier, Isabella Gosch, Jutta Funk, Claudia Luckner-Minden, Patrick Schmidt, Nektarios Valous, Alexandra Tuch, Petra Hill. Auch bei unserer Gruppe im Bioquant möchte ich mich herzlich bedanken: Dr. Niels Halama, Anna Spille, und Anita Heinzelmann. Für das Erfüllen vieler Färbewünsche/Schnittanfertigungen: ein Dankeschön an Tina Lerchl und Rosa Eurich. Dir, Zeynep Koşaloğlu, möchte ich auch herzlich danken für den wissenschaftlichen Austausch, aber auch für die entstandene Freundschaft.

Mein Dank geht auch an Alexander Migdoll, Birgit Eberle, Silke Hirsch und an meine Praktikanten Rebekka Schairer, Daniel Brook, Ines Dhaouadi. Ein weiteres Dankeschön geht an die HSBler und an Ruth Mast, Mariam Salou und Anna-Lena Scherr, denn ohne euch Mädels wäre es nur halb so spannend gewesen.

Ich danke allen beteiligten Kooperationspartnern in diesem Projekt, besonders:

Thomas Hofmann/Ansam Sinjab

Clarissa Gerhäuser/Maria Pudenz

Uwe Warnken

Ein großes Dankeschön geht auch an meine Familie (Mama, Jens und Peter) und Freunde (Chrissy, Steffi, Johannes, Jan, Clara und Flo), die immer ein offenes Ohr hatten und daran geglaubt haben, dass diese Arbeit irgendwann fertig gestellt wird. Hier möchte ich mich ganz herzlich auch bei Verena Kiven für das Korrekturlesen dieser Arbeit bedanken (ich sag nur T-Rex) und für die vielen, gemeinsamen, lustigen Stunden.

Im Besonderen möchte ich mich bei meinem Mann Dirk bedanken, denn egal wie frustrierend es manchmal war, du warst immer da und hast mir Hoffnung und Mut gegeben. Vielen Dank für deine Unterstützung.

Table of contents

Zusammenfassung

Brustkrebs ist die häufigste Krebserkrankung bei Frauen weltweit und führt, trotz guter Therapien, noch oft zum Tode. Obwohl Brustkrebs im frühen Stadium gut therapierbar ist, gibt es bei fortgeschrittener und metastasierter Krankheit nur eingeschränkte Therapiemöglichkeiten.

Das Brustkrebs-assoziierte Antigen NY-BR-1 wurde durch ein serologisches Screening-verfahren (SEREX) entdeckt. NY-BR-1 wird in der Mehrheit der Brusttumoren (>70 %) sowie in Metastasen, in gesundem Brustgewebe, im Hoden und gelegentlich in der Prostata exprimiert. Aufgrund seines begrenzten Expressionsmusters, seiner Immunogenität und seiner Lokalisation im Zytoplamsa und an der Plasmamembran, eignet sich NY-BR-1 als vielversprechendes Target für immuntherapeutische Anwendungen bei Brustkrebspatienten. Das Ziel dieser Doktorarbeit war die molekulare und funktionelle Charakterisierung von NY-BR-1, um die biologische Funktion im normalen sowie im tumorgenen Brustgewebe heraus zu finden.

Um den Effekt der Überexpremierung von NY-BR-1 in zellulären Prozessen, wie Proliferation, Apoptose und Zellzyklus zu untersuchen, wurden HEK293, HEK293T, MCF-10A und MCF-7 mit NY-BR-1 transient transfiziert. Die Ergebnisse im Vergleich mit den Kontrollzellen zeigen, dass NY-BR-1^{+} Zellen nicht proliferieren, nicht vermehrt apoptotisch sind und in der G1 akkumulieren. Diese Erkenntnisse werden durch IHC Färbungen in normalen Brustgewebeschnitten für NY-BR-1 und Ki-67 (Proliferationsmarker) unterstützt, da die Zellen das Expressionsmuster NY-BR-1^{+}/Ki-67^{-} aufweisen.

Eine andere Methode für die Analyse der biologische Funktion von NY-BR-1 war die Identifikation von spezifischen Proteininteraktionspartnern mittels Co-Immunoprezipitation mit NY-BR-1 transfizierten Zelllysaten (HEK293, MCF-10A, MCF-7) mit anschließender Massenspektrometrie. In allen drei Zelllinien konnte das Protein Tubulin 4B-chain identifiziert werden, welches unter anderem in der Mitose eine wichtige Funktion übernimmt. Damit wird verdeutlicht, dass NY-BR-1 eine entscheidende Rolle während des Zellzyklus spielt.

Um die transkriptionelle Regulation vom NY-BR-1 Gen zu untersuchen, wurde eine *in silico* Vorhersage für potentielle Transkriptionsfaktoren Bindungsstellen in der NY-BR-1 Promoterregion mit dem Alibaba 2.0 Algorithmus durchgeführt. Gewebestücke sowie isolierte Epithelzellen aus gesundem Brustgewebe und Tumorzellen aus Pleuralergüssen wurden mit verschiedenen Hormonen, einem Demthylierungsreagenz und Histon-

Deacethylasen Inhibitoren behandelt. Die NY-BR-1 Expression wurde mittels qPCR analysiert. Verschiedene Östrogenrezeptor-Bindungsstellen wurden in der NY-BR-1 Promoterregion und innerhalb des Gens (bis Intron 21) mittels Chip-Seq zugeordnet. Mittles IHC Färbungen konnte verdeutlicht werden, dass das NY-BR-1 Protein und der Östrogenrezeptor hauptsächlich co-exprimiert in normalem Brustgewebe sind. *In silico* Analysen von RNA-Seq Daten konnten zeigen, dass NY-BR-1 vorwiegend in ER^+ Luminal A und B Brustkrebsarten exprimiert wird. Die Stimulation von normalen Brust- sowie Tumorzellen aus Pleuralergüssen mit Östrogen, Tamoxifen oder Progesteron konnte keinen einheitlichen Hoch- oder Herunterregulierungseffekt aufweisen. Auf der anderen Seite konnte eine kurze Stimulation von Tumorzellen aus Pleuralergüssen mit Progesteron, Östrogen, Vitamin D3 und Retinsäure eine Hochregulierung von der NY-BR-1 Expression (8 fach) im Vergleich mit den unbehandelten Kontrollen gezeigt werden. NY-BR-1 Expression konnte auch durch eine VPA Behandlung induziert werden. Es ist erwähnenswert, dass die Kombination von VPA im Vergleich zu den Kontrollen mit den oben beschriebenen Hormonen nur eine schwache Hochregulierung der NY-BR-1 Expression (bis zu 3 fach) zeigen konnte. Diese Resultate legen nahe, dass NY-BR-1 Expression auf Histonlevel und/oder durch Nicht-histon Proteine, wie z.B. Transkriptionsfaktoren, die durch Acetylierung modifiziert werden können, reguliert wird. Die Behandlung der Zellen mit dem Demethylierungsreagens 5´Aza-deoxycytidine konnte keine deutliche Induzierung oder Hochregulierung der NY-BR-1 Expression hervorrufen. Des Weiteren wurde der Methylierungsstatus von ausgewählten Promoterregionen, welche die vorhergesagten CpG Inseln beinhalteten, in bestimmten stimulierten Proben analysiert. Es konnte allerdings keine Assoziation zwischen einem Methylierungsmuster und NY-BR-1 Expression beobachtet werden. Diese Erkenntnis konnte durch eine *in silico* Analyse von Proben aus der TCGA Datenbank bestätigt werden. Auffallend ist, dass eine verringerte NY-BR-1 Expression in kultivierten normalen Brustgewebestücken und isolierten Epithelzellen im Vergleich zu schockgefrorenen Zellen beobachtet werden konnte. Der gegenteilige Effekt war in Brusttumorzellen aus Pleuralergüssen, die in konditioniertem Medium kultiviert wurden, zu beobachten. Eine detaillierte Analyse des konditionierten Mediums wird zeigen, welche Faktoren z.B. Cyotkine, Chemokine, Hormone oder Wachstumsfaktoren, zu einer anhaltenden NY-BR-1 Expression führen.

Unter Beachtung, dass NY-BR-1 Genexpression einem komplexen Regulationsmechanismus unterliegt, dass die Proteinexpression in normalen Brustgewebe mosaikartig ist und dass die Überexpremierung von NY-BR-1 inhibierend in Zelllinien auf Proliferation und G1 Phase im

Zellzyklus wirkt, wurde die Hypothese aufgestellt, dass NY-BR-1 in spezifischen Brustvorläuferzellpopulationen exprimiert wird. Daher wurden Mammosphäroide aus isolierten normalen Brustepithelzellen generiert und analysiert mittels Immunfluoreszenzfärbung und qPCR. In vier analysierten Patienten, konnte die Co-expression von NY-BR-1 in primären Sphäroiden mit Integrin-α6, HER2, GATA-3 und FOXA1 beobachtet werden, wohingegen keine Co-expression mit ER oder PR mRNA detektiert werden konnte. ER/NY-BR-1 IHC Färbungen in normalem Brustgewebe zeigten deutlich, dass sowohl ER^+/NY-$BR\text{-}1^+$ als auch ER^-/NY-$BR\text{-}1^+$ Brustdrüsenzellen vorhanden sind. Diese Resultate lassen darauf schliessen, dass NY-BR-1 in den ER^+/ER^- luminalen Vorläuferzellen in der Brustdrüse exprimiert wird.

Das NY-BR-1 Gen wurde auch hinsichtlich genetischer Variationen untersucht, um Erkenntnisse bezüglich Splice Varianten zu erlangen. Dabei wurden 69 schadhafte („damaging“) nsSNPs innerhalb der kodierenden Region des NY-BR-1 Gens und 39 potentielle splicing SNPs mittels *in silico* Analyse identifiziert.

Zusammengenommen, ist NY-BR-1 hauptsächlich in gut differenzierten Hormonsensitiven Östrogenrezeptor positiven Brustkrebs Subtypen exprimiert und es ist wahrscheinlich, dass NY-BR-1 Expression durch Erα und/oder PR Expression beeinflusst wird. Allerdings muss noch die Assozierung zwischen NY-BR-1 Expression und ER Signalisierung weiter aufgeklärt werden. Desweiteren, wurde bei transienten transfizierten NY-BR-1 expremierenden Zellen eine inhibierende Proliferationsrate und Akkumulation in der G1 Phase zeigt. *In vivo*, endogenes NY-BR-1 wird in nicht proliferirenden (Ki-67 negativ) Zellen in normalen Brustgewebe exprimiert. Das Protein interagiert potentiell mit Tubulin beta-4B chain, welches auf eine entscheidende Rolle in der Mitosis hindeutet. Außerdem konnte gezeigt werden, dass NY-BR-1 in Vorläuferzellen in der Brustdrüse exprimiert wird. Die phenotypische Charakterisierung der Vorläuferzellen und ob das Protein in Krebsstammzellen exprimiert wird, ist Bestandteil laufender Studien.

Zusammenfassend konnte in der vorliegenden Arbeit das erste Mal gezeigt werden, dass NY-BR-1 ein Vorläuferzellmarker in der Brustdrüse ist und den mitotischen Prozess beeinflusst.

Summary

Breast cancer is one of the most common malignancies with increasing incidence every year and a leading cause of death among women. Although early stage breast cancer can be effectively treated, there are limited numbers of treatment options available for patients with advanced and metastatic disease.

The breast cancer associated antigen NY-BR-1 was identified by a serological screening strategy (SEREX). NY-BR-1 is expressed in the majority of breast tumours (>70 %) as well as in metastases, in normal breast tissue, in testis, and occasionally in prostate tissue. Due to its restricted expression pattern, its immunogenicity and its subcellular localization to the cytoplasm and plasma-membrane, NY-BR-1 represents a promising target for immunotherapeutic approaches in breast cancer patients. The aim of this thesis was the molecular and functional characterization of NY-BR-1 to reveal its biological function in normal and tumorigenic breast tissues.

To examine the effect of NY-BR-1 over-expression on cellular processes, such as proliferation, apoptosis and the cell cycle, HEK293, HEK293T, MCF-10A, and MCF-7 were transiently transfected with NY-BR-1. The results showed that NY-BR-1^{+} cells do not proliferate, do not have an increased apoptosis rate and accumulate in the G1 phase compared to the control cells. This finding is supported by IHC stainings of normal breast tissue sections for NY-BR-1 and Ki-67 (proliferation marker) displaying NY-BR-1^{+}/Ki-67^{-} cells.

Another approach to clarify the biological function of NY-BR-1 was the identification of specific protein interaction partners by co-immunoprecipitations with lysates of NY-BR-1 transfected cells (HEK293, MCF-10A, MCF-7) followed by mass spectrometry. The protein identified in all three analysed cell lines is the tubulin beta-4B chain, which amongst other is involved in mitosis, underlining a crucial role of NY-BR-1 during the cell cycle.

To investigate the transcriptional regulation of the NY-BR-1 gene, an *in silico* prediction for potential transcription factor binding sites in the NY-BR-1 promoter region was performed with the Alibaba 2.0 algorithm. Tissue pieces as well as isolated epithelial cells from healthy breast tissue and tumour cells from pleural effusions were treated with different hormones, a demethylation agent and histone deacetylase inhibitors. The NY-BR-1 expression was analysed via qPCR. Several estrogen receptor binding sites were mapped in the NY-BR-1 promoter region and within the gene (up to Intron 21) by ChIP-Seq. The NY-BR-1 protein and the estrogen receptor (ER) are mainly co-expressed in normal breast tissues as assessed by IHC stainings and i*n silico* analyses of RNA-seq data showed that NY-BR-1 is

predominantly expressed in ER positive luminal A and B breast cancers. However, the stimulation of normal breast cells as well as tumour cells from pleural effusions with estrogen, tamoxifen or progesterone did not systematically show an up- or down-regulatory effect. On the other hand, short term stimulations of tumour cells from pleural effusions with progesterone, estrogen, vitamin D3 and retinoic acid lead to an up-regulation of NY-BR-1 expression (up to 8 fold) compared to the untreated controls. VPA treatment also induced NY-BR-1 expression. Of note, the combination of VPA with the above mentioned hormones only showed a weak up-regulation of NY-BR-1 expression (up to 3 fold) compared to the controls. These results suggest that NY-BR-1 expression is further regulated on histone level and/or non-histone proteins such as transcription factors modified by acetylation. Treatment of the cells with the demethylation agent 5´Aza-deoxycytidine did not show a distinct induction or up-regulation of NY-BR-1 expression. Moreover, the methylation status of selected promoter regions containing the predicted CpG islands was assessed in a selection of treated samples, but no association between methylation patterns and NY-BR-1 expression was observed. This finding could be confirmed by an *in silico* analysis of the TCGA database. Strikingly, a decreased NY-BR-1 expression in cultured normal breast tissue pieces and isolated epithelial cells versus snap-frozen cells was seen, which was not observed in breast tumour cells from pleural effusions cultured in conditioned medium. An in depth analysis of the conditioned medium containing the cell free effusion supernatant as well as cell culture experiments with normal breast epithelial cells in conditioned medium will reveal, which factors, e.g. cytokines, chemokines, hormones or growth factors, lead to a sustained NY-BR-1 expression.

Considering the complex regulation of NY-BR-1 gene expression, the mosaic-like protein expression in normal breast tissue and the inhibitory effect of NY-BR-1 over-expression on proliferation and the G1 phase cell cycle arrest in normal cells it was hypothesized, that NY-BR-1 may be expressed by a specific breast progenitor cell population. Thus, mammospheres from isolated normal breast epithelial cells were generated and analysed by immunofluorescent staining and qPCR for NY-BR-1 expression and the presence of progenitor cell markers. In four analysed patients, NY-BR-1 was expressed in primary spheres and co-expression with integrin-α 6, HER2, GATA-3 and FOXA1 could be observed whereas no ER or progesterone receptor (PR) mRNA was detected. IHC staining with ER and NY-BR-1 in normal breast tissue showed that as well ER^+/NY-BR-1^+ as ER^-/NY-BR-1^+ cells in the mammary gland are existing. These results suggest that NY-BR-1 is expressed in the ER^+/ER^- luminal progenitor cells of the mammary gland.

The NY-BR-1 gene was investigated regarding genetical variations to receive a better understanding of splice variants. Thereby, 69 damaging nsSNPs within the coding region of NY-BR-1 gene and 39 potential splicing SNPs could be identified by using *in silico* anlaysis.
Taken together, NY-BR-1 is expressed mainly in well differentiated hormone sensitive estrogen receptor positive breast cancer subtypes and it is likely that NY-BR-1 expression is influenced by ERα and/or PR expression, but the association of NY- BR-1 expression and ER signaling still needs to be elucidated. Furthermore, transiently NY-BR-1 expressing cells show an inhibited proliferation rate and accumulate in the G1 phase. *In vivo*, endogenous NY-BR-1 is expressed in non proliferating (Ki-67 negative) cells in normal breast tissue. The protein potentially interacts with the tubulin beta-4B chain suggesting a crucial role during mitosis. Moreover, NY-BR-1 was shown to be expressed in progenitor cells of the mammary gland. The phenotypic characterization of these progenitor cells and the question whether the protein is also expressed in cancer stem cells is part of ongoing studies. In summary, the presented work could show for the first time that NY-BR-1 is a progenitor cell marker in the mammary gland and is influences the mitotic process.

Abbreviations

% v/v: volume percent
% w/v: weight percent
5'Aza/Aza: 5'Aza-Deoxycytidine
ALDH1: Aldehyde dehydrogenase 1
ANKRD30A: Ankyrin Repeat Domain 30A
AP1: Activator protein 1
APC: Antigen presenting cell
ATP: Adenine triphosphate
BCA assay: Bicinchoninic acid
BCL2: B cell lymphoma 2
BiTE: Bispecific T-cell engager
bp: Base pair
BRCA: Breast cancer 1
bZIP: DNA-binding site followed by leucine zipper motif)
CAR: Chimeric antigen receptor
C/EBP: CCAAT/enhancer-binding protein
CCNE1: Cyclin E1
CD: Cluster of differentiation
CDH1: Cadherin 1
cDNA: Complementary deoxyribonucleic acid
CHR-6494: 3-(1H-indazol-5-yl)-N-propylimidazo[1,2-b]pyridazin-6-amine
CIN: Chromosomal instability
CK: Cytokeratin
CO-IP: CO-immunoprecipitation
C_P: Crossing point
CpG islands: Genomic regions that contain a high frequency of CpG sites
C_T: Threshold cycle
CTL: Cytotoxic T lymphocytes
CTLA-4: Cytotoxic T-lymphocyte associated antigen
dbSNP: Short genetic variations database
DC: Dendritic cell
Dex: Dexamethasone
DMEM: Dulbecco's modified eagle's medium
DMSO: Dimethyl sulfoxide
DNA: Deoxyribonucleic acid
DNMT: DNA methyltransferase
DTT: Dithiothreitol
E2: β-Estradiol
Eb: Elution buffer
ECM: Extracellular matrix
ECL: Enhanced chemi-luminescence
EDTA: Ethylenediaminetetraacetic acid
EGF: Epidermal growth factor
EGFR: Epidermal growth factor receptor
Elf5: E74-like factor 5
ENCODE: Encyclopedia of DNA elements

EpCAM: Epithelial cell adhesion molecule
ER: Estrogen receptor
ErbB2: Erb-b2 receptor tyrosine kinase 2
ERE: Estrogen response element
Erα: Estrogen receptor alpha
ESR1: Estrogen receptor 1
EtBr: Ethidium bromide
EtOH: Ethanol
FACS: Fluorescence-activated cell sorting
FCS: Fetal calf serum
FFPE: Formalin-fixed paraffin embedded
FISH: Fluorescence in situ hybridization
FITC: Fluorescein isothiocyanate
FOXA1: Forkhead box protein A1
FSC: Forward Scatter
GATA3: GATA binding protein 3
GFP: Green fluorescent protein
GGH: Gamma glutamyl hydrolase
GM-CSF: Granulocyte-macrophage colony stimulating factor
GR: Glucocorticoid receptor
GWAS: Genome-wide association study
H3: Histone 3
HAT: Histone acetyltransferase
HBBS: Hank's balanced salt solution
HCL: Hydroclorid acid
HDAC: Histone deacetylases
HER2/neu: Human epidermal growth factor receptor 2
HLA: Human leukocyte antigen
HPV: Human papilloma virus
HuMEC: Human mammary epithelial cell
ICGC: International cancer genome consortium
IDO: Indoleamine 2,3-dioxygenase
IFN-γ: Interferon-gamma
IHC: Immunohistochemistry
IL: Interleukin
kDa: kilo Dalton
LAPTMB4: Lysosome-associated transmembrane protein 4-beta
Lgr5: Leucine-rich repeat-containing G-protein coupled receptor 5
LMD: Laser micro dissection
mAb: Monoclonal antibody
MAPK: Mitogen activated protein kinase
MaSc: Mammary stem cell
MCA: 3-methylcholanthrene
MHC: Major histocompatibility complex
MIC: MHC class I–related chain
min: Minute
mRNA: Messenger ribonucleic acid
MS: Mass spectrometry

MSI: Microsatellite instability
NaCl: Sodium chloride
NCBI SRA: National center for biotechnology sequence read archive
NCT: National Center for Tumor Diseases
NEAA: Non-essential amino acid
NIN: Nucleotide-excision repair instability
NK cell: Natural killer cell
NKG2D: Natural killer group 2, member D
NKR: Natural killer cells receptor
NKT: Natural killer T cells
nm: Nanometer
nsSNP: Non-synonymous SNP
NSEP1: Nuclease sensitive element binding protein 1
NY-BR-1: New York- Breast-1
PAP: Prostatic acid phosphotase
PBS: Phosphate buffered saline
Pen/Strep: Penicillin/Streptomycin
PFA: Paraformaldehyde
PI: Propidium iodide
PIC: Protinase inhibitor cocktail
PR: Progesterone receptor
PSIC: Position-specific independent count
PVDF: Polyvinylidene fluoride
qPCR: Quantitative real time polymerase chain reaction
RANKL: Receptor activator of nuclear factor kappa-B ligand
RAR: Retinoic acid receptor
RNA: Ribonucleic acid
rpm: Rounds per minute
RPMI: Roswell park memorial institute
RT-PCR: Real-time polymerase chain reaction
RXR: Retinoid X receptor
SDS-Page: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEREX: Serological screening strategy
SIFT: Sorting Intolerant from Tolerant
SNP: Single nucleotide polymorphism
SP1: Stimulating protein 1
SSC: Side scatter
sSNP: Synonymous SNP
Stat3: Signal transducer and activator of transcription 3
TAA: Tumor associated antigen
TCGA: The cancer genome atlas
TF: Transcription factor
TGF-β: Transforming growth factor beta
Th: T helper
Thr3: Tyroid hormone receptor
TKI: Tyrosine kinase inhibitor
TNBC: Triple negative breast cancer
TNM: Tumour, Nodes, Metastasis

TP53: Tumour protein p53
TPM: Transcripts per million
Treg: Regulatory T cell
Tris: Tris(hydroxymethyl)aminomethane
TSA: Trichostatin A
TSA: Tumour specific antigen
ULBP/RAET: UL16 binding protein or retinoic acid early transcript
UV: Ultra violet
VEGF: Vascular Endothelial Growth Factor
VitD3: Vitamin D3
VLP: Virus-like particles
V-MYB: Myeloblastosis viral oncogene homolog
VPA: Valproic acid
WHO: World health organization
Wnt: Wingless-type MMTV integration site family
XBP1: X-box binding protein

1 Introduction

1.1 Cancer

Cancer is known since the ancient world and was named by the Greek physician Hippocrates. He was already able to differentiate between carcinoma and sarcoma (Tzortzatou-Stathopoulou, 1994). But it was not possible to diagnose it in a proper way or to cure it in these times. Today, cancer is still an overall issue: the World Health Organization (WHO) cancer report 2014 evaluated that approximately 14 million new cases of cancer and 8.2 million cancer related deaths worldwide could be listed in 2012 (Stewart and Wild, 2014). The problem is that an overall cure of cancer is not within eyeshot even though that in the last decades a lot of research was done to understand the mechanisms behind cancer. But still many questions cannot be fully answered: How does cancer develop? Why some people more succesptible towards developing malignencies? Why do some patients develop resistance towards therapy? Why do some patients get a relapse?

Several theories try to find explanations and one of them is the "multi hit model". It states that multiple somatic mutations in essential genes (oncogenes/tumour suppressor genes) are required to enable uncontrollable cell proliferation. Cell cycle regulatory mechanisms are mainly affected (Lodish, 2008). But it became clear that tumour genesis is a more and more complex process and can differ from patient to patient. It affects different regulatory levels and repair processes of a cell, e.g. epigenetic mechanisms (posttranslational histone modifications, DNA methylation, RNA-based processes (microRNAs)), mutations in the DNA itself or posttranslational modifications. Modifications like single nucleotide polymorphisms (SNPs) or mutations, which can alter protein expression, structure and/or function, can be induced through various ways: virus infections (e.g. human papillomavirus (HPV)), environmental factors (UV light, stress, chemicals), genetic inheritance (nucleotide-excision repair instability (NIN), microsatellite instability (MSI), and chromosomal instability (CIN)), and mobile DNA elements (transposons) (Dunn et al., 2004a). As described in the immunoediting hypothesis also the interplay between cells can adjudicate whether abnormal proliferating cells become a clinical visible tumour or not.

The origin of tumour cells is not fully clarified yet but it is proposed that tumour cell populations have a common ancestor and develop monoclonally at the beginning. It is just due to evolution of a tumour and its genetic instability that the initially homogenous tumour mass becomes heterogeneous and thus, visible and difficult to target for the immune system

(Weinberg, 2014). Furthermore, tumour cells have to develop different strategies to prevent death and hypoxia by ensuring a stable cell homeostasis due to the release of chemokines, cytokines and other signal molecules (e.g. vascular endothelial growth factor (VEGF), transforming growth factor beta (TGF-β)). These factors help to induce angiogenesis towards the growing tumour mass and ensure the important supply of nutrients (Auguste et al., 2005; Holopainen et al., 2011). By secreting a chemical cocktail different immune cells are attracted towards the tumour site and the imbalance of T regulatory (T_{regs}) versus T helper cells can support the tumour outgrowth.

1.1.1 Immunoediting

Over 100 years ago it was proposed that the immune system might play an important role in the fight against cancer. The German physician Paul Ehrlich was the first, who hypothesized that the immune system recognizes tumour cells and is able to fight against these cells (Niederkorn, 2009). Lewis Thomas and Frank Macfarlane Burnet formed the hypothesis of immunosurveillance independent from each other in the late 1950s and they assumed that T cells are crucial in these processes. The immunosurveillance hypothesis describes that the immune system is capable to monitor abnormal proliferating cells, recognizes cell surface structures and activates an elimination process to destroy the mutated cells (Dunn et al., 2002). At this time it was a controversial idea and the hypothesis was criticized by many researchers.

Today, the understanding of cellular processes becomes clearer and therefore the immunosurveillance hypothesis was refined further and is now known as immunoediting. The immunoediting process consists of three stages: elimination (the innate and adaptive immune system destroy all abnormal proliferating cells), equilibrium (the adaptive immune system avoids the tumour outgrowth) and escape (characterizes abnormal cell growth and neither the innate nor the adaptive immune system can hold back tumour growth and metastasis) (Dunn et al., 2002). A tumour's ability to avoid an immune response after a certain time period is depending on its microenvironment. Tumours shape their microenvironment in such a way that immune cells (T-lymphocytes, NK-cells) cannot prevent further outgrowth/tumour genesis and metastasis. The microenvironment itself consists of components varying from patient to patient: cancer and non-cancer cells, secreted soluble factors and non-cellular solid material (including extracellular matrix (ECM)) (Cretu and Brooks, 2007). Thus, the microenvironment drives the immune reaction until the immune system capitulates and the tumour becomes clinically visible. The time period, when and how these changes occur,

varies from patient to patient. The survival strategy of a tumour includes mimicking the immune system to suppress effector cell functions by shaping a specific tumour microenvironment.

1.1.1.1 Elimination Phase

The elimination of tumour cells is carried out by the cooperation between innate and adaptive immunity. Innate immunity is characterized to be an unspecific, not affording long lasting host protection. It is the first defence against invading microbes and bacteria. Adaptive immunity provides the host with long lasting prevention and can be divided into humoral and cellular responses. B- and T-cells and their subpopulations are characteristic for the adaptive immune system (Parham, 2009).

The elimination phase can be divided into a four stage process: it starts with the activation of the innate immune system by the release of danger signals. Macrophages, NK cells, natural killer T (NKT) cells and γδ T-cells recognize the expressed antigens and/or natural killer cells receptor (NKR) ligands of the transformed cells. The danger signals are secreted due to stromal changes including invasive growth of surrounding tissues and starting of angiogenesis (Dunn et al., 2004a, 2004b). The cells of the innate immune system start to produce IFN-γ, which has immunostimulatory and -modulatory effects. The second stage is characterized by a local production of other chemokines to attract more cells of the innate immune system. More macrophages are migrating to the tumour site to secrete cytokines, including interleukin-2 (IL-2), to induce a further production of IFN-γ by NK cells. The release of such a high amount of IFN-γ activates several pathways leading to tumour cell death. In the third stage the adaptive immune response is involved. The dendritic cells (DCs) process and present antigens to activate cellular immune defence (Kushwah and Hu, 2011). These cells are migrating to the transformed cells attracted by the cytokine environment, acquire tumour antigens and after maturation they activate on their part naïve T-cells in lymphoid organs. The activated $CD4^+$ T helper (T_H) 1 cells enhance the development of $CD8^+$ cytotoxic T lymphocytes (CTLs) via antigen cross presentation. In the last step a tumour specific adaptive immunity is acquired to have a long lasting protection against the cancer cells. Thereby the $CD4^+$ and $CD8^+$ T cells migrate to the tumour. $CD8^+$ T-cells induce tumour death via indirect or direct mechanisms affecting apoptosis, cell cycle inhibition and macrophage tumoricidal activity (Dunn et al., 2002, 2004a, 2004b). In the elimination phase the immune system functions as an extrinsic tumour suppressor.

Due to continuous replication of the transformed cells and arising of new neoplastic cells the immune system is constantly required to adapt to new situations. If tumour cells manage to reduce immunogenic signals, the innate immune system gets powerless and the second stage of immunoediting will be entered, with the adaptive immune system controlling the tumour cells.

1.1.1.2 Equilibrium Phase

The equilibrium phase describes a dynamic balance between the adaptive host immune system and the tumour cells, which escaped the elimination phase. The immune system is responsible for holding the tumour cells into dormancy. Therefore, this phase can last over a long period of time depending on the strength of the immune system.

How is the adaptive immune system capable to prevent further proliferation and outgrowth of transformed cells? An explanation refers to evolution: different lymphocytes exercise a selection pressure on genetically unstable tumour cells (Dunn et al., 2004a; MacKie et al., 2003). Therefore a dynamic process is maintained between adaptive immune system and cancer. Over a long time period, tumour cells might change their strategy and may develop resistance towards immune cells due to decreased immunogenicity. The tumour cells do not express any antigens (loss of human leukocyte antigen (HLA) class I, defects in major histocompatibility complex (MHC) genes), which could be recognized by T-cells. But also the immune system can promote further outgrowth of the tumour: if the balance between effector T cells and T_{regs} is disturbed T_{regs} down regulate the T-cells and tumour progression proceeds. $CD4^+$ $CD25^+$ T_{regs} cells inhibit proliferation of other T-cells and the IFN-γ production. A process partly mediated by TGF-β (Mougiakakos et al., 2010; Woo et al., 2001). Different tumour entities (breast, head and neck, pancreas) were analysed with respect to T_{reg} cell accumulation and patient survival. This study found a negative correlation between the amount of T_{reg} cells in the tumour surrounding environment and the overall survival (Liyanage et al., 2002).

The tumour refines through selection pressure and continuous proliferation. When the changes are too advanced the adaptive immune system cannot prevent further progression and the tumour cells escape.

1.1.1.3 Escape Phase

The tumour cells in the escape phase contain a decreased immunogenicity and create their own microenvironment to prevent immune responses. They ensure a self-sustaining supply to grow out and metastasize. The adaptive immune system is no longer able to control the tumour.

The tumour cells release immunosuppressive cytokines such as VEGF, TGF-β or indoleamine 2,3-dioxygenase (IDO) and Chang *et al.* demonstrated that tumours can escape an immune response if they are capable of increasing TGF-β expression or stimulating surrounding cells to produce more TGF-β. The researchers used meth A sarcoma cells to produce clones with an TGF-β overexpression. The effects of overexpression were analysed in mice without a compromised immune system (Chang et al., 1993).

Another way to escape an immune response is to become insensitive for specific cytokines such as IFN-γ. Kaplan *et al.* used mice insensitive to IFN-γ and treated them with MCA to induce tumour development. The mice lacking IFN-γ recognition developed tumours more rapidly and with higher frequency compared to the wild type mice. The researchers also analysed human tumours with a lost ability to bind IFN-γ and suggest that IFN-γ is essential for tumour survival (Kaplan et al., 1998).

Beside the insensitivity against IFN-γ tumour cells have the ability to express additional receptors such as the natural killer group 2, member D (NKG2D) receptor, which has a restricted expression pattern in normal tissues. In an experimental set up 462 patients with primary colorectal cancer were screened and the expression of NKG2D ligands such as MHC class I–related chain (MIC), UL16 binding protein or retinoic acid early transcript (ULBP/RAET) proteins and NK cell infiltration were analysed by tissue microarrays. The results indicated that a selection pressure is on tumour cells which do not express NKG2D. Moreover, high expression of MIC could be associated with NK infiltration and better survival outcome for patients (McGilvray et al., 2009).

The challenging task is now to find targets for developing drugs, which are able to support or reactivate the immune system to fight against cancer.

1.1.2 Epigenetics in cancer

Epigenetic regulation in cancer becomes more and more an important aspect to understand tumour genesis. The term “epigenetic” was introduced by Conrad Waddinton in the 1940s and includes the maintenance of the cellular identity and function (Armstrong, 2014). Epigenetic changes refer to be heritable and reversible in gene expression. These changes can alter the

phenotype but are not encoded in the genotype. DNA methylation, histone modifications, chromatin remodelling, and microRNAs (miRNA) belong to epigenetic regulation mechanisms.

The DNA molecule is organized in chromatin wrapped around nucleosomes, an octamer of two copies of each histone (H3, H4, H2A, and H2B). This kind of organization helps to localize the genetic information into the cell nucleus and to regulate gene expression (Kouzarides, 2007). The histones contain tails, which are positively charged to interact with the negatively charged DNA backbone to form inactive heterochromatin. The histone tails contain a code of methylation, acetylation, phosphorylation, deimination, and ubiquitination and histone composition in conjunction with other non-histone proteins (Cedar and Bergman, 2009; Jenuwein and Allis, 2001). To un-wrap or wrap the chromatin, acetylation is one of the key regulators. Acetyl groups can be added through histone acetyltransferases (HATs) and removed through histone deacetylases (HDACs). After transferring the acetyl groups to histone tails through HATs, their positive charge is neutralized and the interaction between histone tails and DNA backbone is interrupted, leading to a facilitated gene expression. HDACs inhibit gene transcription by removing the acetyl groups, what increases the histone tail/DNA backbone affinity. HDAC inhibitors, such as trichostatin A (TSA) and valproic acid (VPA), target the zinc cofactor at the active site of HDACs and are used in cancer therapy (Arts et al., 2003; Prince et al., 2009).

Beside histone modification, DNA methylation (hypermethylation, hypomethylation, and loss of imprint) is crucial for making a gene accessible for transcription initiation and thus also plays an important role in tumour genesis (Kumar, 2014). In this process a methyl group is added to the 5-carbon position of cytosine (Widschwendter and Jones, 2002). CpG islands are enriched of cytosine and guanine and mostly located upstream of a promoter or exon 1. Expressed genes in normal cells have hypomethylated CpG islands. The opposite takes place in cancer cells, where CpG islands of 5'-end regions are hypermethylated causing gene silencing of e.g. tumour suppressor genes (Bird, 2007; Esteller, 2007). CpG islands can also be found within a gene, the so called gene bodies. The methylation there is associated with repressing transcriptional noise, inhibiting antisense transcription, and relate to replication timing (Aran et al., 2011). In breast cancer, epigenetic mechanisms were observed in two important proteins: ERα and E-cadherin. ERα is encoded by the oestrogen receptor 1 (ESR1) gene and overexpressed mainly in hormone driven breast cancers (luminal A and B). Five CpG islands are present in the promoter region and first exon, respectively. Deregulation of DNA and histone methylation of ESR1 can lead to treatment resistance. E-cadherin is

encoded by the CDH1 gene. It is an adhesion molecule and maintains the differentiated state of mammary gland epithelium. Hypermethylation of CpG islands of CDH1 gene can cause loss of E-cadherin expression in breast cancer and results in an increase of cellular motility with a high risk of invasion of surrounding tissue (Nass et al., 2000). Thus epigenetic regulation needs to be considered to understand the mechanisms of cancer.

1.2 Immunotherapy

The immunotherapy can be defined as a treatment of a disease to induce, enhance or suppress a certain immune response. With cancer immunotherapy, researchers have an additional option to stimulate an immune reaction to destroy cancer cells and direct the equilibrium back to the immune system (Disis, 2014). Depending on the cancer type different approaches can be applied: monoclonal antibodies (mAb) target signal transducing proteins like VEGF (Konno et al., 2010) and human epidermal growth factor receptor 2 (HER2/neu) (Cobleigh et al., 1999; Rueckert et al., 2005), small molecules inhibit for instance kinase domain functions or vaccines immunize the patient against cancer-associated pathogens (e.g. Human papillomavirus (HPV)) (Gillison et al., 2008). Other options of immunotherapy include T cell transfer and specific chimeric antigen receptors (CARs) expressed by T and NK cells to improve therapy outcome.

The basis for cancer immunotherapy is the assumption that the immune system is capable of detecting and destroying abnormally proliferating cells. Today immunotherapy approaches use among others T cell transfer, mAb, small molecules, and different kinds of vaccines to interfere on the tumour itself, on its microenvironment or to start an immune response towards tumour cells. The targets differ depending on the characteristics of the tumour: Tumour cells, immunostimulatory pathways and co-inhibitory pathways can be attacked. The standard cancer therapies are surgery, chemotherapy and radiation but they can cause severe side effects and the therapeutic effect is not always satisfying. In early stages, they can help to reduce the tumour until complete remission. But patients can die because they do not respond towards therapy, become resistant or get a relapse. Another reason for patient mortality are opportunistic infections caused by bacteria, viruses or fungi. Researchers expect from cancer immunotherapy a curative effect with reduced side effects and a support for classical approaches to improve the patients' outcome. Therefore the tumour biology with its microenvironment needs to be understood and new targets need to be defined.

1.2.1 Adoptive Cell transfer and CARs

The general process of adoptive cell transfer includes the isolation of patient specific leukocytes to manipulate them *ex vivo*. These primed cells get injected back into the patient to induce a more effective immune response. The possibilities are not just promising but also versatile.

Autologous T cells can be genetically engineered with so called CARs which recognize a tumour antigen independent of the human leukocyte antigen (HLA) presentation (Rini, 2014). These CARs are composed of a mAb fragment fused to a T cell receptor. Nowadays four generations of CARs are known and they differ in the amount of signalling and co-stimulatory domains. The so called first generation contains one signalling domain, the second one two signalling domains, the third two costimulatory domains and the fourth is primed with additional genes for e.g. pro-proliferative cytokine IL-12 (Brentjens and Curran, 2012).

The CAR T cell approach was tested in different leukaemias and the results are promising. The T_{regs} of the patients were depleted by chemotherapy followed by an autologous CAR T cell transfer, which recognize the surface protein CD19 on B cells. Some patients went into remission of their advanced, progressive B-cell malignancies but some of them experienced severe toxicity due to a cytokine storm (Kochenderfer et al., 2012). Nevertheless, CAR T cells are a promising approach in immunotherapy and clinical trials in different tumour entities are ongoing.

1.2.2 Monoclonal antibodies

The production of mAbs was developed in the 1970s and they are applied for different tumour entities and often in combination with other therapy approaches. These antibodies are developed to target cell surface proteins, which are specifically (over)-expressed in tumours. They block them and prevent further tumour progression. The effect can differ between mAbs depending on their target binding affinity or mechanism of action. In $HER2^+$ breast cancer patients' trastuzumab, also known as Herceptin® (Roche), is used to target the HER2/neu receptor, which is overexpressed in 15-20 % of the breast cancer patients. By binding to the receptor, trastuzumab disrupts the signal cascade of the epidermal growth factor (EGF) and slows down the proliferation of tumour cells (Harries and Smith, 2002). Because patients can react poorly on trastuzumab therapy other mAbs were developed. One of them is pertuzumab, which blocks the HER2/HER3 heterodimerization by binding to the dimerization arm of HER2 (Agus et al., 2002).

The use of mAb in cancer immunotherapy covers a wide range of therapeutic strategies. However, the binding affinity towards a target is insufficient. In huge solid tumours, antibodies might not get equally distributed due to their huge molecular mass of about 150 kDa. The best therapy approach is to use mAb in combination with classical methods.

Another approach for the use of antibodies in immunotherapy is to target and modulate immune checkpoints. These checkpoints refer to inhibitory pathways, which exist in the immune system to prevent autoimmunity but to maintain self-tolerance and homeostasis (Bauzon and Hermiston, 2014). Tumours have the ability to escape the immune system and it is assumed that they trigger specific immune checkpoints to down-regulate T cell activation or effector functions. Several immune checkpoint inhibitors are now in different clinical phases but Ipilimumab (Yervoy, Bristol-Myers Squibb Co, Princeton, NJ), an antibody targeting cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4), is already approved in melanoma (Disis, 2014).

Other mAbs are still situated in different preclinical states, e.g. the mAb P245 to target the transmembrane glycoprotein CD44, which is known to be a breast cancer stem cell marker (Marangoni et al., 2009). The epithelial cell adhesion molecule (EpCAM) is over-expressed in many carcinomas and is now targeted with the mAb MT110 (Solitomab) in a clinical trial phase I. Solitomab is a single-chain bispecific T-cell engager (BiTE) antibody and is tested in different solid tumours (Brischwein et al., 2006; Kwiatkowska-Borowczyk et al., 2015).

1.2.3 Vaccines

Vaccination is a long established method to immunize an organism against against specific pathogens. In cancer immunotherapy, it is one promising approach to use vaccines including whole tumour cell vaccines, tumour lysate vaccines, specific tumour antigens, tumour peptides, heat shock proteins, DNA vaccines, RNA vaccines, dexosomes, and DC-based vaccines (Myc et al., 2011). There are two basic principles in cancer vaccine therapy: on the one hand a preventive treatment to obviate cancer development and on the other hands a curative treatment. If it is used in a therapeutic manner, the patient's tumour needs to express a certain antigen that serves as a target. This antigen is characteristic for the specific tumour. The task of a treatment vaccine is to delay or to stop cancer cell growth, to reduce tumour size by eliminating cancer cells and to prevent a relapse, where other treatments failed before. They have to boost the immune system by increasing the presentation of tumour-associated

antigens (TAAs) to the immune system and need to activate tumour-specific T and B cells (Disis, 2014).

In the recent years mutanome specific vaccination comes into focus. The aim is to design personalized cancer vaccines to improve the therapy and outcome of a single patient. Mutations could serve as novel vaccine targets but unfortunately, just a small number of mutations are shared between the patients and the main amount of mutations is unique in the specific tumour (Stratton, 2011). To translate the individualized therapy approach into the clinic the tumour mutanome needs to be deciphered by next-generation sequencing. The immunogenic mutations are identified by HLA binding predictions or Elispots (Heemskerk et al., 2012). The next step is to generate an individualized mutation-targeted vaccine (Boisguérin et al., 2014). The advantage is that this approach is specially designed for each patient and the possibilities are great to combine different antigens to counteract the selection pressure during tumour evolution (Castle et al., 2012). But this procedure makes it also very cost intensive and it needs to be considered that not every tumour is immunogenic and the heterogeneity between primary tumour and metastases should also be taken in account.

A successful cancer vaccine is sipuleucel-T (Provenge), applied to metastatic, hormone-refractory prostate cancer patients. Autologous antigen presenting cells (APCs) are isolated and *ex vivo* cultured with PA2024 to force the maturation of APCs into mature dendritic cells (DCs). PA2024 is a fusion protein containing prostatic acid phosphatase (PAP), a prostate cancer-associated antigen, and granulocyte-macrophage colony-stimulating factor (GM-CSF). The primed DCs are injected back into the patient and an immune response is induced against PAP^+ prostate cancer cells. Side effects are mainly due to the injection (Higano et al., 2009; Weber, 2014).

An important aspect for the success of vaccines is to determine how the vaccine is administered (intramuscular, subcutaneous, intradermal and mucosal) and which adjuvant is given. One well known vaccine is against HPV infections, which are related to cervix carcinomas. The vaccine contains L1-proteins extracted from the capsid of the four types of papillomaviruses 6, 11, 16 and 18. They cluster spontaneously to virus-like particles (VLP) (No et al., 2011). In some cases the preventive vaccination has to be repeated after 10 years, the therapeutic one even more. Side effects of vaccination can be allergic reactions or inflammation on the site of injection but this is also in some way showing an ongoing immune response. Moreover, it has been reported that after injection of vaccines autoimmune diseases occurred (Amos et al., 2011).

1.2.4 Small molecules

Small molecules have a molecule mass up to 800 Da and can reach every cell. They bind to proteins, polysaccharides, and nucleic acids and can trigger different signal cascades inducing inhibition of tumour progression and growth. In cancer immunotherapy they can target different kinases involved in certain pathways related to cancer incidence. Anderson *et al.* analysed certain small molecule inhibitors affecting the mitogen activated protein kinase (MAPK) pathway. Their results demonstrated that the combination of specific inhibitors resulted in an additive promotion of MAPK pathway by negative feedback release and priming (Anderson et al., 2011). The group of Huertas looked after small molecules connected to specific histones. They demonstrated that the small molecule CHR-6494 reduces phosphorylated histone H3 at Thr3 residue (H3T3ph) levels in a dose-dependent manner. The effect is a mitotic catastrophe characterized by metaphase misalignment, spindle abnormalities and centrosome amplification. This mitotic kinase inhibitor showed antitumor potential in xenografted nude mice without any observed toxicity and brings the cells into G2/M arrest followed by apoptosis (Huertas et al., 2012).

Another approach in immunotherapy settings is the use of tyrosine kinase inhibitors (TKIs). Tyrosine kinases are enzymes and their task is to transfer a phosphate group from ATP to a protein. Due to an additional phosphate a protein can be in- or activated and thus cellular processes can be regulated (Lehninger, 2005). TKIs are applied in solid tumours (e.g. non-small cell lung cancer) (Sorensen et al., 2014) as well as in leukaemia (chronic myeloid) (Brümmendorf et al., 2015). One successful story in different clinical trials is the small molecule TK inhibitor Imatinib (Glivec, Novartis) given to patients diagnosed with gastrointestinal stromal tumours or chronic myeloid leukaemia (Blanke et al., 2008; Christiansson et al., 2015).

The field of application for small molecules in cancer therapy is manifold and the advantages are obvious. They can diffuse to every cell because of their size; they are even capable to passing the blood brain barrier to attack brain tumours. The problem is that small molecules can be distributed throughout the body and the specificity is thus insufficient. They can also bind to healthy cells and might trigger signal cascades with unexpected outcome and severe side effects.

The potential of immunotherapy strategies emerges rapidly but the challenge is still to define new targets to individualize and improve the therapy for every patient.

1.3 Breast cancer

Breast cancer is one of the most common malignancies with increasing incidences every year and a leading cause of death among women. Although early stage breast cancer can be effectively treated, there are limited numbers of treatment options available for patients with advanced and metastatic disease. Most women are diagnosed with breast cancer at the age of 45-65 years. With this tumour entity the female mammary gland tissue and very rarely the male breast tissue are affected. Figure 1.1 represents the anatomy of a female breast (left) with the branching of the glandular tissue. On the right the detailed anatomy of ducts/lobes in pre-pubertal/pubertal condition are shown (A) and branching of ducts and alveolar structures during pregnancy (B).

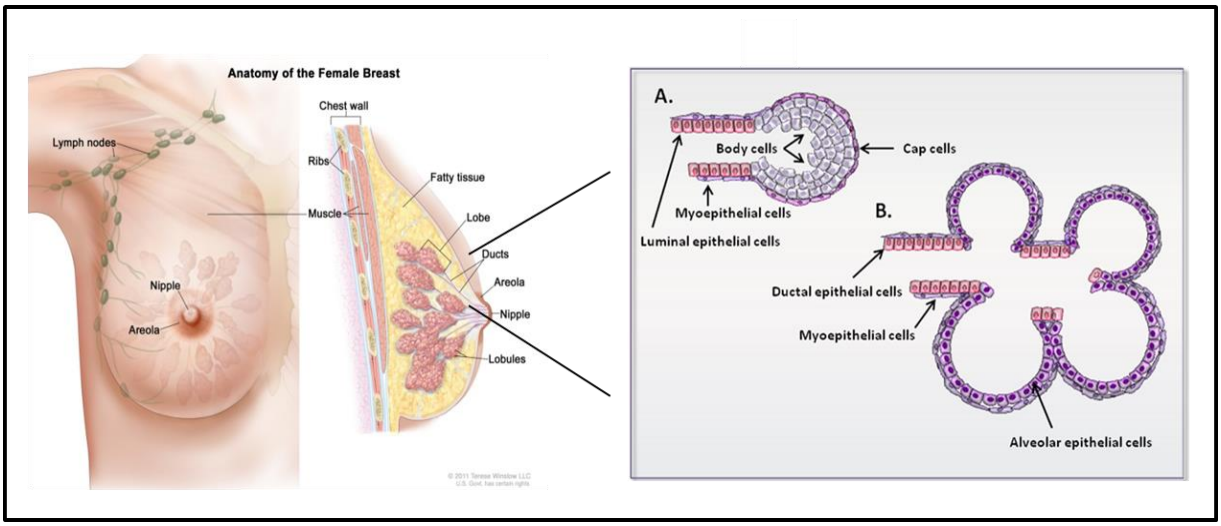


Figure 1.1: Anatomy of the female breast (left) including ductal and lobular structures (right) with (A) the anatomy of ducts/lobes in pre-pubertal/pubertal condition are shown and (B) branching of ducts and alveolar structures during pregnancy. (Figure adapted from the National Cancer Institute, USA; and (Gajewska et al., 2013)*)*

The glandular breast tissue consists of two groups of epithelium: luminal epithelial and myoepithelial cells. The task of myoepithelial cells is to guide the milk through the ductal tree via contractions. The luminal epithelium can be divided into the ductal and alveolar cell lineage whereas the latter produces milk while breast feeding (Fu et al., 2014). The breast is a dynamic tissue and is shaped through different developmental stages in life time such as puberty, adulthood, pregnancy, breast feeding, and menopause. Breast cancer is classified, as any other cancer type as well, into histopathological types, stages and grade of the tumour and expression of specific proteins and genes. Compared to other tumour entities, the breast cancer classification and exact characterization of the tumour is of great importance because therapy strategies are very specialized, e.g. hormone sensitive (ER/PR) versus triple negative

tumours require different therapies. Breast cancer can be subdivided into four groups: $HER2^+$, Luminal A or B, and triple negative. Depending on the molecular subtype, condition of the patient (health, age), and stage/grade/progression of the tumour, different treatment options are available such as surgery, chemo-, radiation-, hormone-, and targeted therapy. Table 1.1 lists which targeted therapy options are available for the different molecular subtype.

Table 1.1: Summary of treatment options for breast cancer patients according to their molecular subtype (Goldhirsch et al., 2013).

Breast cancer type	**Receptor status**	**Treatment options**
$HER2^+$	Her2 over-expressed or amplified ER^-/PR^-	Cytotoxics and anti-HER2 (Trastuzumab, Pertuzumab, Ertumaxomab, and Lapatinib)
Luminal A	$ER^+/PR^+/HER2^-$ Ki-67 'low'	Endocrine therapy (Tamoxifen)
Luminal B	**Luminal B-like ($HER2^-$)** $ER^+/HER2^-$ and at least one of: Ki-67 'high' PR 'negative or low' **Luminal B-like ($HER2^+$)** ER^+/HER2 over-expressed or amplified Any Ki-67 Any PR	Endocrine therapy for all patients, cytotoxic therapy for most (Tamoxifen) Cytotoxics and anti-HER2 and endocrine therapy (Tamoxifen)
Triple negative/Basal-like	Triple negative (ductal) $ER^-/PR^-/HER2^-$	Cytotoxics

As with all tumours, common risk factors are alcoholic abuse, low or no exercise, obesity respectively a high carb/fat diet, and genetic predisposition (BRCA1/BRCA2 mutations). Hormones are the key drivers in shaping the breast during life time. Therefore, additional risk factors for getting breast cancer can be hormones in contraceptives or hormones given in menopause, hormonal therapy, menstruation at early age, and dense, glandular breast tissue.

It is assumed that pregnancy can prevent the outgrowth of hormonal driven cancer and it is of note that an early pregnancy is more preventive compared to a late one. A study compared parous versus nulliparous glands in mice and the results showed that the Wnt:Notch signalling ratio in the basal stem/progenitor cells was changed (Meier-Abt et al., 2013). Human breast

epithelial cells are under hormonal influence, which also shows consequences on the epigenome (Pal et al., 2013).

The molecular subtypes express different genes/receptors and therefore respond differently to therapy. 50-60 % of breast cancer patients belong to the luminal A subtype and the tumours are characterized by a low histological grade, low degree of nuclear pleomorphism, and low mitotic activity (thus low levels of proliferation related genes). Beside ER expression and genes associated with ER regulation (hepatocyte nuclear factor 3 alpha (FOXA1), X-box binding protein 1 (XBP1), GATA binding protein 3 (GATA-3), B cell lymphoma 2 (BCL2), erbB3 and erbB4, the luminal A tumours also express luminal epithelial cytokeratin's (CK) 8 and 18. Patients diagnosed with luminal A have a good prognosis and a lower relapse rate compared to other subtypes.

Compared to luminal A, luminal B tumours are more aggressive, more proliferative and have a higher histological grading. 15-20 % of breast cancers belong to this subtype. Patients have a worse prognosis, a higher recurrence rate, and become quicker insensitive to endocrine therapy. The luminal B tumours have a distinct gene expression pattern: they mainly express growth receptor signalling genes and proliferation-associated genes like myeloblastosis viral oncogene homolog (v-MYB), gamma glutamyl hydrolase (GGH), lysosome-associated transmembrane protein 4-beta (LAPTMB4), nuclease sensitive element binding protein 1 (NSEP1) and cyclin E1 (CCNE1). The distinction between luminal A and B is difficult and therefore the Ki67 value, a proliferation marker, was established. This marker was introduced by Cheang *et al.*, who set the cut-off point of Ki67 for luminal A tumours on 14 % and everything above is considered as luminal B. Unfortunately, the use of Ki67 marker is not standardized and thus it has limitations depending on the used antibody, and how it is analysed (Cheang et al., 2009).

The HER2 gene, also known as ERBB2, encodes for a member of the EGF receptor family of receptor tyrosine kinases. 15-20 % of breast cancer patients carry the HER2 subtype. These tumours are characterized by a high expression of HER2 and genes associated with the HER2 pathway, high proliferation rate and an aggressive behaviour, which refers to a higher tumour grading. Around 50 % of $HER2^+$ tumours do express low levels of ER. The uncertainty of HER2 positivity on immune-histochemical level is circumvented by using the fluorescence-in-situ-hybridization (FISH) technique. In general, patients with $HER2^+$ tumours have a poor prognosis due to hormonal agent resistance and metastases in brain and visceral organs but can be treated inter alia with the monoclonal antibodies such as trastuzumab alone or in combination with pertuzumab (Yersal and Barutca, 2014).

The characteristic features of triple negative tumours are that they are infiltrating ductal tumours, highly aggressive, have a high histological and nuclear grade, and give rise to metastasis in brain and lung. The term “triple negative” is used in clinical setting and refers to no expression of ER, PR, and HER2. The term “basal-like” originates from gene expression microarray analysis. These tumours express myoepithelial markers such as CK5, CK 14, CK 17 and laminin and over-express P-cadherin, fascin, caveolins 1 and 2, αβ crystallin and epidermal growth factor receptor (EGFR). This tumour type is insufficient in DNA repair mechanisms and DNA damage sensing (DNA double strand breaks). 8-37 % of breast cancer patients receive the triple negative/basal-like diagnosis. The triple negative type can be divided into several subtypes - one of them is grouped as breast cancer 1 (BRCA1) positive. Mutations in the caretaker gene BRCA1 are frequently found in basal like tumours and characterized by an early relapse rate and metastasis formation. They do express abnormal basal cytokeratins, TP53, EGFR, and P-cadherin. Another subtype is referred to as claudin-low and its main features are a low expression of tight junction gene and cell-cell adhesion (claudin 3, 4, and 7) genes but a high expression of stem cell features. The prognosis for all triple negative cancers is poor (Yersal and Barutca, 2014).

Adjuvant and neo-adjuvant therapy is administered to breast cancer patients but this is depending on the molecular subtype and health condition of the patient (Miller et al., 2014).

The challenge in breast cancer is to carefully characterize the heterogeneous tumour mass to overcome relapse and resistance and find new targets for immunotherapy.

1.4 Stem cells in the female breast

In the recent years stem cells have become an important issue because of their potential to differentiate into any specialized cell type. They can give rise to more stem cells by undergoing mitosis and are able to replenish adult tissues. Breast tissue contains several different multipotent stem/progenitor cells. But in contrast to other stem cell systems, (e.g. hematopoietic) breast stem cells are difficult to isolate and to expand without losing their characteristic features. It is of note that gene expression profiles of mammary stem cells (MaScs) can undergo dramatic changes during pregnancy, lactation, and menopause. Moreover, it seems that the adult MaScs differ strongly from fetal MaScs due to the hormonal influence during puberty. MaScs numbers vary between the myoepithelial/basal compartment and luminal epithelial cells and are enriched in the myoepithelial/basal compartment. It seems that the MaScs population is heterogeneous containing subsets either expressing leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5) or not (Visvader and Stingl, 2014).

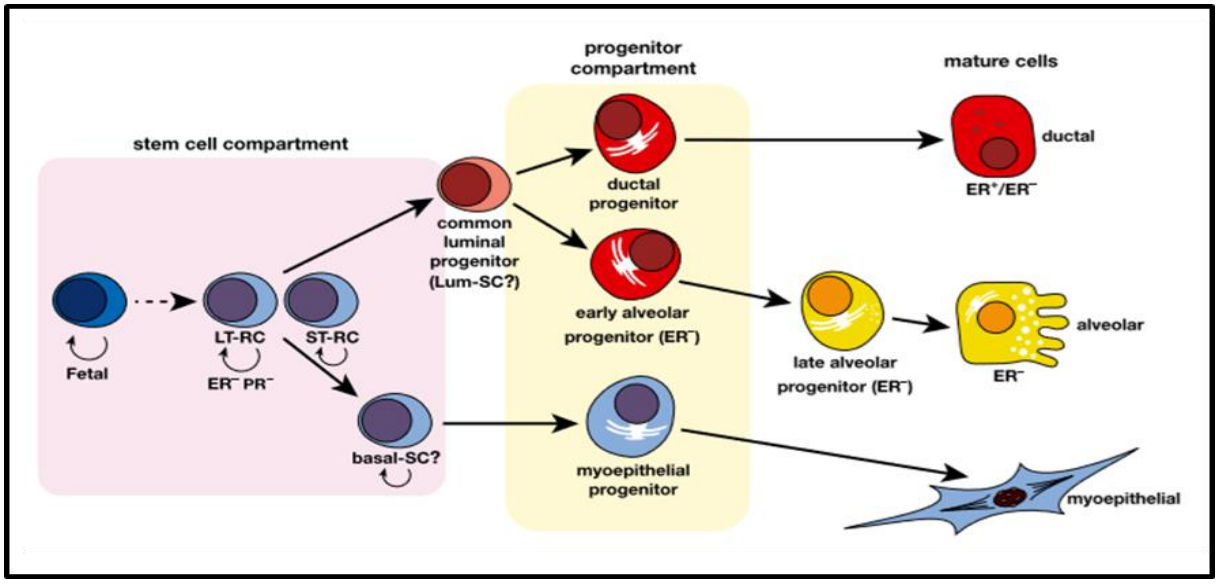


Figure 1.2: Stem and progenitor cells in breast tissue with some specific expression features (adapted from (Visvader and Stingl, 2014)*).*

Epithelial cells with high proliferative potential are mainly found in the ducts and with less frequently in the lobules, which holds true for luminal and myoepithelial cells. The different stages of potential stem and progenitor cells in the breast tissue can be viewed in figure 1.2.
The MaSc have a low activity of aldehyde dehydrogenase 1 (ALDH1) and do not express ER, PR and ErbB2. By paracrine signalling of ER^+/PR^+ neighbouring cells, MaScs can be forced to proliferate. In contrast, the luminal progenitors ($CD49f^+/EpCAM^+$) consisting of ER+ as well as ER- cells and have a high ALDH1 activity (Eirew et al., 2012). Moreover, the luminal progenitors are under control of Elf5, Stat3, GATA-3, C/EBPβ and BRCA1 depending on the stage of development (Makarem et al., 2013). Elf5 expression can be stimulated through progesterone via receptor activator of nuclear factor kappa-B ligand (RANKL) signalling and is responsible for basal to luminal cell fate decisions. RANKL is a regulator of MaScs activity and as GATA-3 it forces luminal progenitors into alveologenesis during pregnancy. In summary the luminal progenitors are responsible for morphogenesis during puberty and alveologenesis. After involution, parts of the progenitor cells will diminish via apoptosis and a new pool is recruited for another round of alveologenesis.
The myoepithelial progenitor cells are characterized to be $CD24^{low}$ and expressing the basal marker CK14 and the adhesion molecule CD49f, also known as integrin alpha 6. But due to the low amount it is difficult to segregate all the different progenitor cells from each other.

In table 1.2 is an overview of the current markers for isolating stem and progenitor cells from human breast tissue. The marker panel of each subtype depends on how the cells were cultured, expanded and analysed.

It is assumed, that the different molecular subtypes in breast cancer arise due to mutations and modifications in the different subsets of stem/progenitor cells of the female breast (Al-Hajj et al., 2003). Therefore, it is more and more important to characterize the gene and protein expression profile of progenitor cells for a better understanding of breast cancer.

Table 1.2: Overview of characteristic features of each human stem/progenitor cell subtype (adapted from (Oakes et al., 2014)*).*

Marker	Multipotent stem cell	Luminal lineage	Myoepithelial lineage
ALDH1	Breast stem/progenitor cells $ALDH1^+$	Differentiated human luminal cells ($EpCAM^+/CD49f^+/ALDH1^-$); Undifferentiated human progenitor luminal cells ($EpCAM^+/CD49f^+/ALDH1^+$)	/
CD10			Is expressed
CD24	Depending on study: $CD24^{-/low}$	$CD24^{+/high}$	$CD24^{high}$
CD44	Tumor initiating cell $CD44^+/CD24^{-/low}$	/	/
CD49f		$CD49f^{neg}/EpCAM^+$ (mature luminal) $CD49f^+/EpCAM^+$ (luminal progenitor)	$CD49f^{high}$
cKIT	/	$cKIT^{high}$	
EpCAM	$CD49f^+/EpCAM^{high}$ or $EpCAM^{low}/CD49f^{bright}$	$EpCAM^{high}$	$CD49f^+/EpCAM^{-/low}$
Sca1		$CD24^{high}/Sca1^+$	
Thy1/CD90	Bipotent progenitor $EpCAM^+/CD49f^+/MUC1^-/CD133^-$ $(CD10/Thy1)^+$	Luminal restricted progenitor $EpCAM^+/CD49f^+/MUC1^-/CD133^-$ $(CD10/Thy1)^+$	$MUC1^-/CD133^-$ $(CD10/Thy1)^+$

1.5 The breast cancer associated antigen NY-BR-1

The breast cancer associated antigen New York-Breast-1 (NY-BR-1) was identified by a serological cloning strategy (SEREX) by Jäger *et al*. (Jäger et al., 2001). NY-BR-1, also known as ANKRD30A, is located on chromosome 10p11-p12. There are several transcripts

annotated in the Ensembl database. Due to alternative splicing processes transcripts are known containing between 36 and 42 exons. For transient transfection experiments the full length variant (aa 1-1397) was used, which corresponds to the Ensembl ID ENST00000611781. Computational analyses revealed several functional protein motives, including five tandem ankyrin motifs, a bZIP site (DNA-binding site followed by leucine zipper motif) and a bipartite nuclear localization signal motif. These structural features suggest on the one hand that NY-BR-1 might be involved in protein-protein interactions and on the other hand NY-BR-1 might function as a potential transcription factor (Jäger et al., 2001). But up to date, the functional aspects of this 158 kDa protein are still unknown.

NY-BR-1 is expressed in the majority (>70 %) of breast tumours as well as metastases, in normal breast tissue, in testis and occasionally in prostate tissue. Biochemical labeling and transient expression studies showed that NY-BR-1 encodes a novel membrane protein, which can be detected by a monoclonal antibody on the cell surface of transfected cells (Seil et al., 2007). Furthermore, NY-BR-1 was confirmed as an immunogenic antigen since NY-BR-1-specific spontaneous humoral immune responses in 7 % of breast cancer patients could be detected (Jäger et al., 2001; Seil et al., 2007; Varga et al., 2006). Theurillat *et al.* performed an *in situ* hybridization experiment for exon 4-7 and 30-33 and NY-BR-1 mRNA detection was stronger for exons 30-33 than for exons 4-7. From these results one can conclude that splice variants of NY-BR-1 might exist lacking 5' sequences (Theurillat et al., 2008). Theurillat *et al.* also mapped putative oestrogen responsive elements (ERE) as binding sites for ERα and binding sites for stimulating protein 1 (SP1) as well as activator protein 1 (AP1) in the NY-BR-1 promoter region by *in silico* analysis, suggesting the transcriptional regulation of NY-BR-1 by ERα. Another analysis showed that NY-BR-1 mRNA expression correlates with ERα protein expression. In a cohort study the researcher analysed NY-BR-1 expression in combination with oestrogen receptor alpha (ERα), HER2, EGFR, HLA class I antigen expression and overall survival. They could show that patients with strong NY-BR-1 expression had a better prognosis compared to patients with NY-BR-1 negative tumours. NY-BR-1 expression correlated with ERα and HLA class I antigen expression but not with HER2 or EGFR (Theurillat et al., 2007).

Thus, due to its restricted expression pattern and over-expression in the majority of breast tumours as well as its immunogenicity, NY-BR-1 represents a promising target antigen for immunotherapeutic approaches and may also serve as a target for antibody based therapies.

2 Aims of the project

The latest studies show that breast cancer is still the second most common cancer in the world and the most common one among women. As described before (see 1.3) different kinds of therapies are available but due to resistance towards therapy and recurrence of cancer new targets need to be defined. NY-BR-1 would be an interesting target for immunotherapy for several reasons depicted in chapter 1.5 but up to date the biological function, regulatory mechanisms and interaction partners of NY-BR-1 in healthy and tumorous tissue is completely unknown.

The aim of this project was to characterize this protein on the molecular and functional as well as genetic level and therefor the following questions were addressed: I) How is the NY-BR-1 expression transcriptionally regulated? II) What is the effect of NY-BR-1 over-expression on cellular processes such as proliferation and cell cycle? III) Are there specific protein interaction partners? IV) Is NY-BR-1 a progenitor cell marker in the mammary gland? And V) Does an integrative *in silico* analysis of the NY-BR-1 gene reveal interesting SNPs or differences in methylation on certain promoter fragments, which might alter transcription initiation?

To investigate the transcriptional regulation of the NY-BR-1 gene, the promoter region was analysed for putative transcription factor binding sites with the Alibaba 2.1 tool. This algorithm predicted numerous binding sites for nuclear hormone receptors (oestrogen receptor alpha (ERα), progesterone receptor (PR), glucocorticoid receptor (GR) and retinoid X receptor (RXR)/retinoic acid receptor (RAR) in the assumed promoter region of ANKRD30A (figure 2.1) based on sequence comparison. To study the hormonal transcriptional activation *in vitro*, several NY-BR-1 negative or weakly positive breast cancer cell lines (e.g. T47D, 226L, ZR-75-1, MCF-7), tissue pieces from breast reductions, isolated epithelial cells and pleural effusions from breast cancer patients were stimulated with different hormones, the demethylation agent 5'Aza-Deoxycytidine and the histone de-acetylation inhibitors VPA and/or TSA. NY-BR-1 expression was analysed by qPCR and the methylation status via EpiTyper MassARRAY technology.

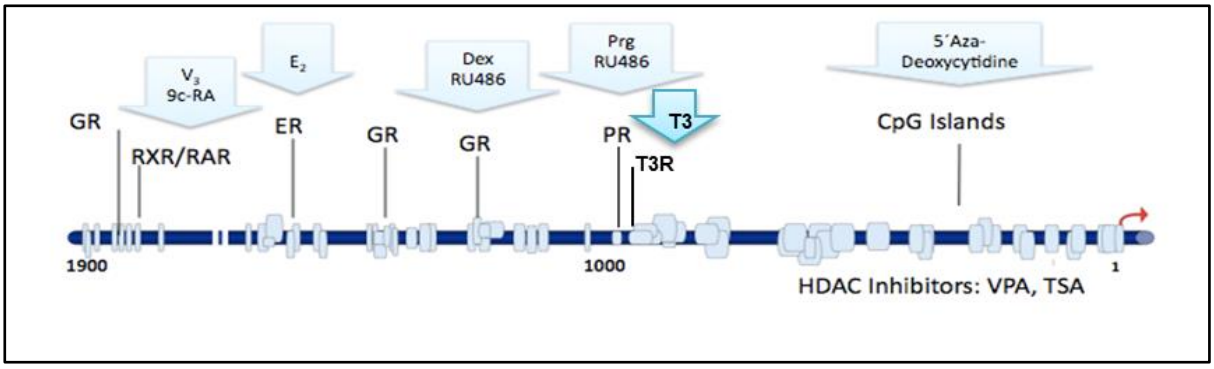


Figure 2.1: Predicted binding sites for nuclear hormone receptors and CpG islands in the promoter region of NY-BR-1. (CpG Islands: genomic regions that contain a high frequency of CpG sites; HDAC Inhibitors: Histone deacetylase inhibitors; VPA: Valproic acid; TSA: Trichostatin A; and T3R: Thyroid hormone receptor)

To examine the effect of NY-BR-1 over-expression on cellular processes, proliferation assays and cell cycle analyses by fluorescence-activated cell sorting (FACS) analysis were performed with transiently transfected HEK293, HEK293T, MCF-10A, and MCF-7 cell lines. Immunohistochemical analyses of mammary gland tissue derived from healthy patients show a mosaic-like NY-BR-1 expression pattern. Single luminal epithelial cells in a mammary gland are NY-BR-1 positive. This characteristic expression pattern can be due to expression on a specific progenitor cell population. To further evaluate this assumption, mammospheres were generated. The spheres were analysed by immunofluorescent staining and qPCR was performed to characterize the gene expression profile in more detail. Interaction partners can give hints in which pathways the protein of interest might be involved. To identify specific protein interaction partners of NY-BR-1, co-immune precipitations were performed with lysates of NY-BR-1 transfected cells (HEK293, MCF-10A, MCF-7) and lysates of normal breast/breast tumour tissue. Western Blot and silver staining were used to detect proteins after electrophoretic separation on polyacrylamide gels. Finally, an analysis of the proteins was done by mass spectrometry.

Furthermore, an extensive analysis utilizing genomics, transcriptomics and epigenetics was performed with publicly available data from databases such as dbSNP, ICGC, ENCODE, and NCBI SRA.

Considering our results, illustrated in the next chapters, we hypothesized that NY-BR-1 is a potential progenitor cell marker in breast tissue. Additionally, we suggest that NY-BR-1 is expressed in specific molecular subtypes of breast cancer patient, namely the ones which are hormone sensitive. It is likely that NY-BR-1 expression is influenced by ERα and/or PR expression.

3 Material and Methods

3.1 Materials

3.1.1 Chemicals

- 4-Hydroxytamoxifen (Sigma, H7904)
- 5'Azadeoxycytidine (Aza) (Sigma, A3656)
- Dexamethasone (Dex) (Sigma, D4902)
- β-Estrodiol (E2) (Sigma, E2257)
- Progesterone (Sigma, P0130)
- 9-cis Retinoic acid (Sigma, R4643)
- Trichostatin A (TSA) (Sigma, T1952)
- Vitamin D3 (Vit D3) (Sigma, D1530)
- Valproic acid (VPA) (Sigma, P4543)
- Silver nitrate (Sigma, 20139)
- Ethanol (EtOH) (Roth, 9065.2)
- Methanol (MeOH) (Roth, 4627.2)
- Acetic acid (Roth, 52016)
- Ethidium bromide (EtBr) (Roth, 2218.1)
- Saponin (Roth, 4185.1)
- Potassium tetrathionate (Sigma, P2926)
- Potassium acetate (Grüssing, 12240)
- Potassium carbonate (Roth, P743.1)
- Sodium thiosulfate-pentahydrate (Sigma, 21,726-3)
- Tris(hydroxymethyl)aminomethane (Tris) (Roth, 5429.3)
- Hydrochlorid acid (HCl) (Roth, 46251)
- Formalaldehyde [37 %] (Roth, 4979.1)

3.1.2 Antibodies

- α-NY-BR-1#2: monoclonal mouse (Jäger et al., 2007)
- α-NY-BR-1 (C-18): polyclonal goat (SantaCruz Biotech, sc-163686)
- α-Integrin alpha 6 (CD49f): monoclonal mouse (abcam, ab20142)

- α-Estrogen Receptor alpha: monoclonal rabbit (abcam, ab32063)
- α-E-Catenin: polyclonal rabbit (Cell signaling, 3236)
- E-cadherin (24E10): monoclonal rabbit (Cell signaling, 3195)
- Keratin 17/19 (D4G2) XP®: monoclonal rabbit (Cell signaling, 12434)
- EpCAM (D1B3): monoclonal rabbit (Cell signaling, 2626)
- α-human Ki67: monoclonal mouse (MIB1, Dako, M7240)
- Progesterone receptor A/B (D8Q2J) XP®: monoclonal rabbit (Cell signaling, 8757)
- Purified Goat IgG (invitrogen™, 02-6202)
- Purified Mouse IgG1 κ Isotype Control (BD, 554121)
- Rabbit (DA1E) mAb IgG XP® Isotype Control (Cell signaling, 3900)

- Goat α-mouse IgG Pox conjugated (Dianova, 115-035-003)
- Donkey α-goat IgG-HRP (Santa Cruz, sc-2056)
- Donkey α-rabbit IgG-HRP (Santa Cruz, sc-2077)
- Donkey α-goat IgG H&L (FITC): polyclonal (abcam, ab7121)
- Goat α-mouse IgG (FITC) (Dianova, 115-096-062)
- Donkey α-goat IgG (FITC) (abcam, ab7121)
- Goat α-rabbit (FITC) (Dianova, 111-096-045)
- Goat α-mouse (FITC) (Dianova, 111-096-045)
- Goat α-rabbit (Cy3) (Dianova, 115-166-045)
- Mouse EpCAM-FITC (BD, 347197)
- HER2 – FITC (BD, 340553)
- mouse IgG1-PE Isotype Control (BD, 550953)
- Mouse IgG1-FITC Isotype Control (BD, 555909)
- mouse α-human Cytokeratin 14, 15, 16, and 19 PE (BD, 550953)

3.1.3 Dye

- APC Annexin V (BD Pharmingen™, 550475)
- 7-AAD (BD Pharmingen™, 559925)
- PKH26 (Sigma, 081M1013)
- Propodium iodide (PI) (Sigma, P4170)
- DAB Kit (Leica, DS9800)

3.2 Methods

3.2.1 Primary material

Healthy mammary tissues from breast reduction were obtained from Clinic for Plastic and Reconstructive Surgery, Aesthetic and Preventive Medicine, Heidelberg University Hospital-Ethianum, Heidelberg, Germany. The pleural effusions were provided from the department of Gynecology and Obstetrics, National Center for Tumor Diseases (NCT), Heidelberg University Hospital, according to the ethic vote number S-069/2010. Paraffin embedded breast tumour tissue sections were kindly provided by the National Center for Tumor Diseases (NCT) Tissue Bank, Heidelberg in accordance to the ethic vote number S 206/05.
Table 3.1 shows the obtained primary material used for several experiments.

Table 3.1: Overview of the obtained primary material. HD-T-176 – HD-T-283: Tissue pieces; HD-T-292 – HD-T-311: Isolated epithelial cells; HD-A-58 – HD-A-88: Ascites/pleural effusion cells

Sample ID	Material
HD-T-176	Healthy breast tissue
HD-T-177	Healthy breast tissue
HD-T-178	Healthy breast tissue
HD-T-180	Healthy breast tissue
HD-T-203	Healthy breast tissue
HD-T-258	Healthy breast tissue
HD-T-283	Healthy breast tissue
HD-T-292	Healthy breast tissue
HD-T-311	Healthy breast tissue
HD-A-58	Ascites of breast cancer patient
HD-A-66	Pleural effusion of breast cancer patient
HD-A-68	Ascites of breast cancer patient
HD-A-73	Ascites of breast cancer patient
HD-A-78	Ascites of breast cancer patient
HD-A-86	Ascites of breast cancer patient
HD-A-88/HD-A-90	Ascites of breast cancer patient

3.2.2 Cell culture methods

3.2.2.1 Cell culture

Cell lines were cultured with the specific medium described in table 3.2 at 37 °C, 5 % CO_2. The medium was refreshed every three to four days. Dulbecco's Modified Eagle's Medium (DMEM) (DMEM/F12) and Roswell Park Memorial Institute (RPMI) medium were purchased from Sigma (D0819; D8062; R2405).

Table 3.2: Overview of the used cell lines, the type of cells and their specific culture medium.

Cell line	Type	Culture medium
HEK293	Human embryonic kidney (ATCC-Reference: CRL-1573)	DMEM with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
HEK293T	Human kidney with SV40 large T antigen (ATCC-Reference: CRL-11268)	DMEM with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
MCF-7	Mammary gland/breast derived from pleural effusion (ATCC-Reference: HTB-22)	RPMI with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
MCF-10A	Mammary gland/breast (ATCC-Reference: CRL-10317)	DMEM/F12 with 10 % FCS, 1 % Pen/Strep, 5 µg/ml Insulin, 10 µg/ml Hydrocortisone, 8 µl/500ml Cholera Toxin, 20 ng/ml hEGF)
KS	Breast carcinoma (Kayser et al., 2003)	RPMI with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
ZR-75-1	Mammary gland/breast, derived from ascites (ATCC-Reference: CRL-1500)	RPMI with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
T-47D	Mammary gland; derived from pleural effusion (ATCC-Reference: HTB-133)	RPMI with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
SK-BR-3	Mammary gland/breast; derived from pleural effusion (ATCC-Reference: HTB-30)	DMEM with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA
MDA-MB-453	Mammary gland/breast; derived from pericardial effusion (ATCC-Reference: HTB-131)	DMEM with 10 % FCS, 1 % Pen/Strep, 1 % Hepes, 1 % NEAA

3.2.2.2 Transfection

Cell lines were transfected according to the manufacturer´s protocol either with lipofectamine (Invitrogen™, Lipofectamine® 2000 Transfection Reagent, 11668-019) or with X-tremeGENE HP DNA Transfection Reagent (Roche, 06366236001). The ratio between DNA and Transfection reagent was always 1:3. Cells in 6 well plates and in T-25 culture flasks were transfected with 4 µg DNA, cells in T-75 culture flasks with 20 µg DNA. However, 4 µg pc3 GFP was used in T-75 culture flasks due to achieve comparable GFP fluorescence to the NY-BR-1-GFP fusion protein. Before adding the DNA-transfection mixture to the cells, the culture medium was removed and replaced by DMEM or RPMI without any additives.

3.2.2.3 Cell counting

For counting the cells were diluted 1:2 with trypan blue, a vital stain to separate dead from living cells. Cell counting was done using a hemocytometer.

3.2.2.4 Pleural effusion or ascites regeneration

Pleural effusion or ascites are accumulations of fluids either close to the lungs or the peritoneal cavity. They can contain blood and various other cells like tumour cells and lymphocytes. Due to potential infectious material, before starting, all equipment (collection container, drain) was carefully desinfected. The material was filled into sterile bottles, then divided into 50 ml falcon tubes and centrifuged for 10 min at 1500 rpm, at room temperature. Supernatant was removed, filled into new falcon tubes and centrifuged again for 10 min at 3500 rpm, at room temperature. The cleared supernatant was later used to produce conditioned medium (supernatant mixed with RPMI 1:2). The cell pellets were brought together, resuspended in RPMI medium without any additives and centrifuged for 10 min at 1500 rpm, at room temperature. The supernatant was discarded and RPMI without additives was added. The cells were seeded into T75 culture flasks with a density of $2{,}5x10^6$/ml – $5x10^6$/ml in max. 8 ml RPMI and incubated between 1 – 1.5 hours at 37 °C, 5 % CO_2, to allow macrophages to stick to the flask ground. The macrophage-free supernatant was removed containing only the non-adherent cells and tumour cells respectively and placed into new falcon tubes. The cells in the culture flasks were washed three times with PBS, which was then added to the non-adherent cells. Conditioned medium was added to the macrophages and they were cultured at 37 °C, 5 % CO_2. The non–adherent cells were centrifuged for 10 min at 1500 rpm, at room temperature. They were counted with trypan blue using a hemocytometer. Depending on the amount of cells and the following experiments, the cells were seeded into T25/T75 culture flasks with a density of $2.5x10^6$/ml or in 6 well plates with a density of $1x10^6$/ml.

3.2.2.5 Isolation of mammary gland epithelial cells

Skin and fat were removed from the supposed glandular tissue, which was stored in DMEM/F12 medium until the preparation of the whole tissue was finished. 50 ml falcon tubes were prepared containing 70 % EtOH for the first wash and DMEM/F12 for the second and third wash. The tissue pieces were dipped quickly three times in each wash solution by holding them with sterile pincers. They were left in wash solution three. Few tissue pieces were transferred to a big petri dish without medium and were cut into small pieces with fine scissors. Before the minced glands were transferred into new falcon tubes for further digestion, they were checked for big organoids and then equally distributed among the falcons.

The composition of the digestion mix was 1 x collagenase (Roche, 10103586001), hyaluronidase (Sigma, H3506-1G), 1 x glutamine (Life technologies, 25030123), 5 % FCS and DMEM/F12 medium with a total volume of 20 – 30 ml/falcon. Collagenase A was dissolved in HBSS (Gibco®, 24020-091) and filtered through a 0.22 µM filter. Hyaluronidase (stock: 1 mg/ml (8X)) was freshly prepared because its activity lasts for just 10 min. It was dissolved in HBSS and sterile filtered using a 0.22 uM filter. 1.25 ml of the 1 mg/ml solution was added to 10 ml digestion mix. The falcon tubes were sealed with parafilm and incubated overnight at 37 °C with gentle shaking (40 rpm). After digestion a fat layer, a medium layer and at the bottom the organoid layer was observed. The falcons were centrifuged for 2 min at 80 g and 4 °C. Carefully the supernatant was removed starting with the fat layer until the organoid layer was reached. The organoids were washed with 1x HBSS and centrifuged for 5 min at 350 g and 4 °C. The supernatant was taken off and the second enzymatic digestion was performed. 0.25 % Trypsin – EDTA (Life technologies, 25300096) was pre-warmed and depending on the organoid amount 10-15 ml were added and resuspended for 3 to 5 min or until the organoids got smaller in size. The enzymatic reaction was stopped with 30 ml cold HF (1x HBBS, 2 % FCS) and then centrifuged for 5 min at 350 g and 4 °C. The supernatant was removed carefully and if erythrocytes were present 0.2 % NaCl solution was added for 3 to 5 min. The reaction was stopped by adding HBSS and then centrifuged for 5 min at 350 g and 4 °C.

Dispase (Roche, 4942078001) was dissolved in modified HBSS (HBSS + 1 % Hepes) to get a 5 mg/ml concentration and sterile filtered using a 0.22 µM filter. 10 ml of pre-warmed dispase was used for each pellet. DNAse (stock: 10 mg/ml) (Roche, 11284932001) was filtered using 0.22 µM filter and added in a 1:15 ratio (DNAse:Dispase). After adding the dispase/DNase mix the organoids were resuspended to fully dissociate them. The enzymatic reaction was stopped with 30 ml cold HF (1x HBBS, 2 % FCS). On top of a 50 ml falcon tube a 40 µM mesh (BD, 352340) was placed and wet with 2 ml HBSS. The digested organoids were filtered by pipetting slowly and gently over the mesh to avoid clogging of the filter. The cells were centrifuged for 5 min at 350 g and 4 °C. The supernatant was removed carefully and the cells were washed with 1x HBSS to remove residual FCS. They were centrifuged again for 5 min at 350 g and 4 °C. The cell pellet was resuspended in plating medium (HuMEC medium with 5 µg/ml Insulin (Sigma, I9278), 4µg/ml Heparin (Sigma, H3149), 100 U/ml Penicillin (Sigma, P4458), 100 µg/ml Streptomycin (Sigma, P4458), 0.5 µg/ml Hydrocortisone (Sigma, H0888), 20 ng/ml hEGF (Sigma, E9644), 20 ng/ml bFGF (Life technologies, PHG0021), and 1x B27 (Life technologies, 17504044)) or frozen (45 % medium, 50 % FCS, 5 % DMSO).

3.2.3 Functional assays

3.2.3.1 Proliferation Assay

To examine the effect of NY-BR-1 over-expression or down-regulation on cellular processes in different cell lines (HEK293, HEK293T, MCF-10A, and MCF-7), cells were transiently transfected as described before with 20 µg of pcDNA3-NY-BR-1-GFP plasmid to discriminate NY-BR-1 expressing cells from non-transfected cells. A pcDNA3-GFP plasmid and non-transfected cells served as controls. The cells were cultured at 37 °C, 5 % CO_2 in serum free medium. 24 hours after transfection the cells were labeled with PKH26 (proliferation marker) according to the manufacturer's instructions, and were incubated for one-five days.

3.2.3.2 Apoptosis and viability assay

Shortly before FACS analysis, the cells were stained with 7-AAD (live/dead staining) and Annexin V APC (apoptosis marker) at each time point according to the manufacturer´s instructions.

3.2.3.3 Cell cycle analysis

Cell cycle analysis is done to reveal the information about cell ploidy, cell position in the cell cycle and the amount of apoptotic cells arising from fractionated DNA.

To analyse the effect of NY-BR-1 over-expression on the cell cycle the following assay was performed. The cell lines (HEK293, HEK293T, MCF-10A, and MCF-7) were transiently transfected with either 20 µg of pcDNA3-NY-BR-1-GFP or 20 µg of pcDNA3-GFP (as a control) 24 hours prior to the cell cycle analysis. The second control was composed of non–transfected cells treated the same way as transfected cells. The supernatant was taken off containing the dead cells and placed into falcon tubes. All three cell populations were harvested with a cell scraper, added to the appropriate supernatant and centrifuged for 10 min at 1500 rpm. The supernatant was discarded and the cell pellets were washed with 30 ml PBS. After counting a minimum of $1x10^6$ cells/population the cells were used for cell cycle analysis. This time point is referred to as day one. The rest of the cells were seeded into 6 well plates for cell cycle analysis on day two and three. While vortexing the cell pellet 5 ml ice cold 70 % ethanol was added drop by drop. For a minimum of two hours the cells were stored at -20 °C. The cells were washed twice with PBS/2 % FCS, resuspended in 425 µl PBS and brought to a FACS tube. RNase A (Qiagen, 19101) was added to the cells with a final concentration of 100 µg/ml and incubated for 30 min at 37 °C in the dark. PI (live/dead

marker) was administered with a final concentration of 100 µg/ml. The cells (min. 50.000) were slowly (max. 300 cells/sec) measured in a FACS CantoII.

3.2.3.4 Intra and extracellular antibody staining

Ascites cells were analysed by FACS using the α-NY-BR-1#2 (mouse) antibody for intracellular (IC) and extracellular (EC) staining as well as EpCam (EC), CK 14/15/16/19 (IC) and HER2 (EC) antibodies. EpCam/HER2/CK14/15/16/19 antibodies were used according to the manufacturer´s instructions. The equivalent isotypes were used as a control.

For an intracellular (IC) staining, the cells were fixed with 2% PFA for 15 min at room temperature, whereas the cells for extracellular (EC) staining were kept on ice. The fixed cells were centrifuged for 5 min at 1500 rpm and permeabilized with 0.5 % saponin buffer (PBS, 5 % FCS, 0,5 % BSA, 0,1 % - 0,5 % saponin) for 15 min at room temperature. The cells were centrifuged for 5 min, 1500 rpm and the primary antibodies, diluted in either saponin buffer (IC) or FACS buffer (PBS, 5 % FCS, 0.1 % NaN_3) (EC), were added to cells for IC and EC staining for 30 min at room temperature (IC) or ice (EC). The α-NY-BR-1#2 (Mouse) antibody was diluted to a final concentration of 26 µg/ml, α-NY-BR-1 (C-18) to 4 µg/ml, the mouse IgG1 isotype to 26 µg/ml and the goat IgG isotype to 4 µg/ml.

The cells were washed after adding 1 ml saponin buffer (IC) or FACS buffer (EC) for 5 min at 1500 rpm. The secondary antibodies, diluted in either saponin buffer (IC) or FACS buffer (EC), were administered and incubated for 20 min at room temperature (IC) or ice (EC). The final concentrations were 30 µg/ml for α-mouse conjugated with FITC of and 20 µg/ml for α-goat FITC. The cells were washed with either saponin buffer or FACS buffer for 5 min, 1500 rpm and were analysed via FACS CantoII.

3.2.3.5 Flow cytometry

FACS is a method to analyse single cells. The method is based on the specific light scattering and fluorescent pattern of each cell (Jaye et al., 2012).

The cells were analyzed at a FACS CantoII (BD Biosciences, Heidelberg) with the software FACS DIVA (Version 6.1.2, BD Biosciences, Heidelberg). The FACS CantoII system contains three different lasers: a blue (488-nm), a red (633-nm) and a violet (405-nm) laser. Auto-fluorescence cells were taken as a control to determine the individual auto-fluorescence of a cell population. By combinations of multiple fluorescence marker single labeled cells were used for compensation. The cells were sorted at first according their size (forward scatter (FSC)) and granularity (side scatter (SSC)). The adjacent gate settings depend on the analysis.

Proliferation assay: A gate was set in the FSC/SSC application to get all cells for further analysis but leaving the cell debris out. The next application was to discriminate GFP positive from GFP negative cells. Therefore, the PE-A (area) vs. FITC-A channel was adjusted. The PKH26 label beams in the PE channel, the GFP transfected cells in the FITC channel. To analyse the proliferation profile of each cell population a histogram was used with the PE-A setting. The last step was to discriminate the apoptotic cells from living/dead cells and for this a quad gate was set in the Annexin APC-A/7-AAD-A setting.

Cell cycle analysis: No gate was set in the FSC/SSC application. The next step was to discriminate NY-BR-1 GFP positive from NY-BR-1 GFP negative cells. This was done using the FITC-A channel and a histogram. In both cell population (NY-BR-1 $GFP^{+/-}$) a gate was set in the PI-A vs. PI-W (width) setting to gate the doublets out. The cell cycle was analysed in the PI-A histogram setting on the basis of the previous gate setting.

Intracellular and surface staining: A gate was set in the FSC/SSC application to get all the tumour cells for further analysis but leaving lymphocytes and cell debris out. In a histogram, using the FITC-A setting or PE-A setting, the appropriate isotpye control was taken and the appendant staining.

The final analysis and graphical representation was performed using the software FlowJo (Version 7.6.5, Tree Star, Ashland, USA).

3.2.3.6 Stimulation Assay

To analyse a hormone dependent transcriptional regulation of NY-BR-1, healthy mammary gland tissues of 3 to 5 mm size, isolated epithelial cells from healthy breast tissues or pleural effusions were used. The mammary gland-containing tissue pieces were incubated with phenolred-free DMEM/F12 medium, the isolated epithelial cells with HuMEC basal medium (Gibco, 12753-018) and the pleural effusion cells with conditioned medium (cleared supernatant of pleural effusion mixed 1:2 with serum free RPMI medium).

The healthy mammary gland tissues and the different kinds of cells were treated with different hormones/histone deacetylase (HDAC) inhibitors/reagents (see Table 3.2.3) alone or in combination. The stimulation time was depending on the agent, which was used. 5‘Azadeoxycytodine was added each day to the cells for four to six days, and all the other substances were added for six-seven hours. After stimulation was done the tissue pieces were collected with a needle, frozen in liquid nitrogen and stored at -80 °C. The cells were harvested with a cell scraper, washed twice with PBS to remove the medium, the cell pellet was divided for either RNA or DNA isolation and stored at -80 °C.

Table 3.3: Overview of the hormones, HDAC inhibitors and reagents used for up- and down regulation of NY-BR-1 in stimulation experiments.

Hormones	HDAC inhibitors	Reagents
10 nM E2	1 mM VPA	100 nM 4-Hydroxytamoxifen
1 µM Progesterone	200 nM TSA	1 µM 5‘Azadeoxycytodine
1 nM Vit D3		
1 µM 9-cis retinoic acid		
1 µM Dexamethasone		
1 nM T3		

Table 3.4: Overview of the used agent for the stimulation experiments and their mechanism of action.

Agent	Function
5’Aza	• Causes DNA demethylation or hemi-demethylation • DNA demethylation can regulate gene expression by "opening" the chromatin structure detectable as increased nuclease sensitivity • Allows transcription factors to bind to the promoter regions, assembly of the transcription complex, and gene expression (www.sigmaaldrich.com)
Dex	• Glucocorticoid anti-inflammatory agent, acts through GR • Regulates T cell survival, growth, and differentiation • Inhibits the induction of nitric oxide synthase (www.sigmaaldrich.com)
E2	• .Major estrogen secreted by the premenopausal ovary, acts through ER • Estrogens direct the development of the female phenotype by regulating gene transcription • It also induces the production of gonadotropins which, in turn, induce ovulation. • Exposure to estradiol increases breast cancer incidence and proliferation (www.sigmaaldrich.com)
VitD3	• Vitamin D acts through its VDR that is a member of the ligand-dependent transcription factor superfamily • Modulates the proliferation and differentiation of both normal and cancer cells. • Has antiproliferative and antimetastatic effects on breast cells (www.sigmaaldrich.com)
Progesterone	• Endogenous steroid hormone, acts through the different kind of PRs • Its effect is amplified by oestrogens
Retinoic acid (9cis)	• Ligand for both the retinoic acid receptor (RAR) and the retinoid X receptor (RXR) that act as transcription factors to regulate the growth and differentiation of normal and malignant cells (www.sigmaaldrich.com)
VPA	• Histone deacetylase inhibitor
TSA	• Histone deacetylase inhibitor
Tamoxifen	• Tamoxifen is a selective estrogen response modifier (SERM), protein kinase C inhibitor and anti-angiogenetic factor • Tamoxifen is a prodrug that is metabolized to active metabolites 4-

	hydroxytamoxifen (4-OHT) and endoxifen by cytochrome P450 isoforms CYP2D6 and CYP3A4 • In breast cancer, the gene repressor activity of tamoxifen against ERBB2 is dependent upon PAX2 • Blocks estradiol-stimulated VEGF production in breast tumor cells (www.sigmaaldrich.com)

3.2.4 Mammospheres

3.2.4.1 Generation of mammospheres

Mammospheres have the ability to renew and differentiate into cells with unipotent and bipotent potential (Dontu et al., 2003). To generate mammospheres, the cells were resuspended at a density of 1x10^5 cells/ml and plated in a 10 cm plate (Corning Petri Dishes, Ultra low binding surface, Sigma, CLS3262-20EA) with 10 ml plating medium. The plate was swirled to avoid having cells attaching to the periphery and incubated at 37 °C, 5 % CO_2. If secondary or tertiary spheres are required more cells need to be plated. Four to five days after plating 5 ml of plating medium was added, Seven to nine days after plating the mammospheres were collected in cap filters of FACS tubes. It is of note that the pipettes and cap filters should be pre-wetted with medium, otherwise the spheres will stick to it. The ultra-low culture dishes were washed with more medium to collect remaining spheres, and filtered through the mesh again. The cap filters were tapped on a plate cover to remove remaining medium and then flipped over a 15 ml falcon tube. The spheres were eluted by slowly adding medium with releasing the pressure every once in a while to avoid bubble formation from below. The spheres were centrifuged for 5 min at 350 g and 4 °C. The supernatant was removed and 1 ml pre-warmed Accumax (Sigma, A7089) was added. The spheres were resuspended until all spheres were dissociated. To stop the reaction medium was added and the cells were centrifuged for 5 min at 350 g and 4 °C. The cells were resuspended in freshly prepared HuMEC with additives, and were plated by pooling in half of the initial volume. To generate tertiary spheres the process was repeated after one week of incubation at 37 °C, 5 % CO_2.

3.2.4.2 Immunofluorescence of mammospheres

The spheres were collected in cap filters of FACS tubes. They were washed 3 times with 1X PBS, fixed with 4 % PFA in PBS for one hour at room temperature, washed 3 times with 1 x PBS and permeabilized with 0.1 % Triton-X100 in PBS for 5 min at room temperature. Blocking was done with 1 % BSA in 0.1 % Triton-X100 in PBS for 30 min at room temperature. The primary antibodies were diluted in blocking solution and added to spheres in

filters. The filters were sealed before with parafilm to avoid leaking. The incubation lasted for one hour at room temperature. The spheres were washed 3 times with 1 x PBS and incubated with the secondary antibodies FITC-coupled goat anti-mouse or Cy3-coupled goat anti-rabbit (1:450 in PBS with Hoechst, 1:1000 w/v, to detect DNA) for 45 min at room temperature. Hoechst was added to the secondary antibody solution (1:1000 w/v in PBS) to detect DNA. Finally, the spheres were collected with 50 µl PBS and transferred to slides. PBS was allowed to dry for 30 min to one hour in the dark, before adding Aquatex® (Merck Millipore, 1085620050) and covering with rectangular coverslips. During and after incubation with secondary antibodies the slides were protected from light. The cells were examined using a confocal laser scanning microscope (IX81, Olympus) with an x20, x40 and x60 oil objective. Images were collected and processed using the FV10- ASW 2.0 Viewer software (Olympus) and ImageJ.

The staining was performed in the Research Group Cellular Senescence of Dr. Thomas Hofmann, German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, Heidelberg, Germany.

3.2.5 Molecular techniques

3.2.5.1 Bacteria (E.coli) preparation

The XL1 blue bacteria (endA1 gyrA96 (nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)) were obtained from StrataGene (Heidelberg). They were cultured in Lysogene Broth-Medium (LB - medium), pH 7.5 consisting of pepton 10 g/L, yeast extract 5 g/L, NaCl 10 g/L over night at 180 rpm and 37°C. Depending on the plasmid the bacteria contained 100 µg/ml carbenicillin (Roth, 6344.2) or 40 µg/ml chloramphenicol (Roth, 3886.2) was added to the LB - medium.

3.2.5.2 Plasmids

Table 3.2.5 summarizes the used plasmids for transfection and transient expression of the encoded proteins in different assays.

Table 3.5: Overview of used plasmid constructs in several experiments. Plasmid numbers 58/65 are NY-BR-1 full length constructs (Seil et al., 2007) *and 42/126 are the empty pcDNA3.1 vector.*

Filemaker No.	**Construct**	**NY-BR-1 amino acid**	**Size of encoded protein [kDa]**
58	pc3-NY-BR-1	1 - 1397	154
65	pc3-NY-BR-1-GFP	1 - 1397	181
42	pcDNA3.1 (-)		
126	pcDNA3.1 (+)		
99	pcDNA3.1 (+)-GFP		Derivative with extended cloning site and a C-terminal "Green fluorescent protein" (GFP) sequence (Rosorius et al., 1999)

3.2.5.3 RNA Isolation

RNA isolation was performed according to the manufacturer's protocol (Qiagen; RNeasy Plus Mini Kit, 74134). An additional DNase digestion was done according to the manufacturer's protocol to remove genomic DNA (Invitrogen™; DNase I, Amplification Grade; 18068-015).

3.2.5.4 Genomic DNA Isolation

DNA isolation was performed according to the manufacturer's instructions (Promega; Wizard SV Genomic DNA Purification System, A2360)

3.2.5.5 Genomic DNA Isolation with bisulfit conversion

DNA isolation was performed according to the manufacturer's protocol (Zymoresearch; EZ DNA Methylation-Direct™ Kit; D5020).

3.2.5.6 Plasmid-DNA Isolation

Plasmid-DNA Isolation was done with either Plasmid Miniprep Kit (Qiagen, 27104) or with Plasmid Midiprep Kit (Qiagen, 12145) according to the manufacturer's protocol.

3.2.5.7 Photometric determination of DNA/RNA concentration

DNA/RNA concentration was measured with nanodrop (Nanodrop ND-1000, Thermo Scientific). Either aqua dest. or elution buffer (EB buffer (10 mM Tris-Cl, pH 8.5), Qiagen,

19086) served as blank. 1 µl/sample was used and the extinction was measured at 260 nm and 280 nm. The ratio of the wavelengths 260 nm/280 nm shows the purity of the sample: a value of 1.8 for DNA 1.8 and 2.0 for RNA testifies great purity.

3.2.5.8 DNA Sequencing

To sequence plasmids or DNA sequences the service of GATC Biotech (Konstanz, Germany) was used. The sequences were analysed with the programmes BioEdit (Version 7.053, Tom Hall) and MegAlign (LaserGene Suite, Version 10.1, DNAStar, USA).

3.2.5.9 Reverse transcriptase for cDNA synthesis

For cDNA 150 ng RNA was used and was performed according to the manufacturer's protocol either with (Invitrogen™; SuperScript® III Reverse Transcriptase; 18080-044) or (Thermo Scientific; Maxima H Minus First Strand cDNA Synthesis Kit; K1651).

3.2.5.10 Generation of whole cell lysates

Medium was removed and the cells were washed with cold PBS. Cells were harvested with a cell scraper in cold PBS and centrifuged for 5 min at 1300 rpm and 4 °C. After removing the PBS the cells were lysated with Co-IP buffer (50 mM Tris-HCl, pH 8; 5 mM EDTA, pH 8; 150 mM NaCl; 0.5 % NP-40) containing 1 mM DTT and 1 % protease inhibitor cocktail (PIC) (Roche, 10261000) for 20 min on ice. The lysate was centrifuged for 10 min at 14.000 rpm and 4 °C to get the proteins separated from the rest of the cells. The supernatant containing the proteins was taken off and filled into a new tube. The lysates were kept on ice all the time and later stored at -80 °C.

3.2.5.11 Protein quantification

Protein analysis was performed with the BCA assay according to the manufacturer's protocol (BCA Protein Assay Kit, Thermo Scientific, 23227). Lysates were diluted 1:5 and 1:10 with aqua dest. The amount of protein was determined with a photometer (BioPhotometer plus, eppendorf).

3.2.5.12 Oligo nucleotides

All oligo nucleotides were purchased by Biospring AG, Frankfurt. They were created with the program Seqbuilder (LaserGene Suite, Version 10.1, DNAStar, USA). The parameter of each was calculated with OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html)

and the specificity was controlled with the tool Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Table 3.6: Oligos used for expression analysis with filemaker number of the AG Jäger laboratory; name of the gene of interest; forward/reverse sequence, melting temperature and size of the product.

Filemaker number	Gene	Oligo-Sequence	TM [°C]	Size[bp]
1 2	ß-Actin	FW: 5'-GGCATCGTG ATGGACTCC G-3' RV: 5'-GCTGGAAGGTGGACAGCG A-3'	55.4 55.4	631
21 22	NY-BR-1	FW: 5'-CAAAGCAGAGCCTCCCGAGAAG-3' RV: 5'-CCTATGCTGCTCTTCGATTCTTCC-3'	58.6 57.4	931
23 24	NY-Eso-1	FW: 5'-CAGGGCTGAATGGATGCTGCAGA-3' RV: 5'-GCGCCTCTGCCCTGAGGGAGG-3'	58.8 64.1	332
182 183	SSX2	FW: 5'- GGATCCATGAACGGAGACGACGCCTTTGC-3' RV: 5'- GTCGACCTCGTCATCTTCCTCAGGGTC-3'	65.7 64.3	642
67 68	ß-Actin	FW: 5´-GCGGGAAATCGTGCGTGACATT-3´ RV: 5´-GATGGAGTTGAAGGTAGTTTCGTG-3´	56.7 55.7	232
110 109	ERα	FW: 5'-AGGCTGCGGCGTTCGGC-3' RV:5'-AGCCATACTTCCCTTGTCAT-3'	56.7 49.7	272
113 111	ERβ	FW: 5'-TTCCCAGCAATGTCACTAACT-3' RV: 5'-TCTCTGTCTCCGCACAAGG-3'	50.5 53.2	527
115 114	pS2	FW: 5'-TGGAGCAGAGAGGAGGCAA-3' RV: 5'-GCCGAGCTCTGGGACTAATCA-3'	53.2 56.3	347
241 242	Maspin	FW: 5'-TGCTGCCTACTTTGTTGGCAAGT-3' RV: 5'-TGATACTGTCAATGTTTCCCATACAGA-3'	55.3 55.2	149
249 250	VDR	FW: 5'-ATGGCGGCCAGCACTTCCCTGCCTGAC-3' RV: 5'-CTCCTCCTTCCGCTTCAGGATCATCT-3'	67.3 61.1	330
259 258	p21	FW: 5'-CAGGAGGCCCGTGAGCGATGGA-3' RV: 5'-TCAGCCGGCGTTTGGAGTGGTAGA-3'	62.3 60.8	346
262 263	Cyp24	FW: 5'-CTCATGCTAAATACCCAGGTG-3' RV: 5'-TCGCTGGCAAAACGCGATGGG-3'	52.4 58.3	300
264 265	RXRα	FW: 5'-CGACCCTGTCACCAACATTTGC -3' RV: 5'-GAGCAGCTCATTCCAGCCTGCC-3'	56.7 60.4	142
276 277	GR	FW: 5'-TGGCTGTCGCTTCTCAATCAGACT-3' RV: 5'-ACACAGCAGTGGATGCTGAACTC-3'	57.4 57.1	300
286 287	PR	FW: 5'-CCGACGCCGAGCCCAAGGACGAC-3' RV: 5'-CGACCCCGAGGAGGACGCAGACGA-3'	66.0 65.9	275
402 403	Sgk1	FW: 5'-CCTTGGGCTACCTGCATTCACTG-3' RV: 5'-CCCGAGCCGCTTTGTCCTG-3'	58.8 57.6	398
880 881	YY1	FW: 5'-ACCTGGCATTGACCTCTCAG-3' RV: 5'-TCTCCGGTATGGATTCGCAC-3'	53.8 53.8	345
883 882	JunB	FW: 5'-ACTTTTCTGGTCAGGGCTCG-3' RV: 5'-GTAGCTGCTGAGGTTGGTGT-3'	53.8 53.8	362
885 884	CEBPα	FW: 5'-AGACGTCCATCGACATCAGC-3' RV: 5'-TTGCTGTTCTTGTCCACCGA-3'	53.8 51.8	673
887 886	C1orf61	FW: 5'-CTCCGTGTGGACGATAGAGC-3' RV: 5'-CTCCCAACTTAGCCATGCCA-3'	55.9 53.8	303
904 905	c-Jun	FW: 5'-GTCCGAGAGCGGACCTTATG-3' RV: 5'-CTTTTTCGGCACTTGGAGGC-3'	55.9 53.8	756
909 908	SP1	FW: 5'-GCCTGTAGCGCCCTTTGTGG-3' RV: 5'-TCCTGGCCTGGGCTGGTAGAGAA-3'	57.9 60.6	417
911 910	Max1	FW: 5'-GGAGAGCGACGCTGACAAA-3' RV: 5'-TTCTTCCTGCTTTGGGGCTC-3'	53.2 53.8	408
913 914	CTCF	FW: 5'-TGCCGTTACTGTGATGCTGT-3' RV: 5'-TAGCTGTTGGCTGGTTCTGT-3'	51.8 51.8	604
919 920	c-Myc	FW: 5'-GTGGTCTTCCCCTACCCTCT-3' RV: 5'-GCTGCGTAGTTGTGCTGATG-3'	55,9 53.8	356
921 922	THRα	FW: 5'-GCCGCACAATCCAGAAGAAC-3' RV: 5'-CCAATGTCATCGGGCAGGAA-3'	53.8 53.8	358
60 59	NY-BR-1	FW: 5'-CAAGAGCTCTGCAGTGTGAGATTG-3' RV: 5'-CTGGTATTGGTGTTTCAGTGTGGC-3'	57.4 57.4	315
100	NY-BR-1	FW: 5'-GAGGTGCTCCATCAACCACTTTC-3'	57.1	456

80		RV: 5'-TCATGAGTTTTCTGTTTCTGCTTTC-3'	52.8	
225 226	ABCG2	FW: 5'-TGGAGATTCCACTGCTGTGGCA-3' RV: 5'-TGACCTGCTGCTATGGCCAGTG-3'	56.7 58.6	690
943 944	GATA-3	FW: 5'-GAACCGGCCCCTCATTAAG-3' RV: 5'-GTTAAACGAGCTGTTCTTGGG-3'	53.2 52.4	288
923 924	Integrin-α 6	FW: 5'-GCTAAACCTTCCCAGGTGTATT-3' RV: 5'-AGACTCCGTTAGGTTCAGGGA-3'	53.0 54.4	274
931 932	FoxA1	FW: 5'-AGGGCTGGATGGTTGTATTG-3' RV: 5'-TGGCATAGGACATGTTGAAGG-3'	51.8 52.4	245
256 257	ERBB2	FW: 5'-TGACCTGCTGGAAAAGGGGGAGCG-3' RV: 5'-TCCCTGGCCATGCGGGAGCATTCAG-3'	62.5 64.2	150

3.2.5.13 Real-time Polymerase chain reaction (RT-PCR)

In the RT-PCR defined DNA sequences are amplified. Amongst others, it can be used to check the presence of genomic DNA within RNA samples. The following protocol was performed with the GoTaq® Flexi DNA polymerase (Promega, M8305) (see Table 3.2.6) and PCR reactions were carried out in a T3000-„Cycler" (Biometra, Jena).

Table 3.7: Master mix per sample using the GoTaq® polymerase.

Reagents	**PCR**
H2O	13,3 µl
5x Green or Colorless GoTaq®Flexi Buffer	5 µl
25 mM MgCl2	3 µl
10 mM each d'NTP Mix	2 µl
10 µM Oligo	Each 0.5 µl
GoTaq®G2 Flexi DNA Polymerase (5u/µl)	0.2 µl
Template (RNA/cDNA)	1 µl

*Table 3.8: Protocol to amplify specific sequences in RT-PCR. * Time for annealing was calculated due to oligo length and its GC-amount, time for elongation was depending on the size of the specific gene part.*

Step	**Temperature**	**Time**	**Cycle**
Initial denaturation	94-96 °C	60-120 s	
Denaturation	94-96 °C	30-60 s	35
Annealing *	T_M-5 °C	20-60 s	
Elongation *	70-72 °C	20-300 s	
Final elongation	72 °C	60-180 s	
Cooling	16 °C	Until remove	

3.2.5.14 Quantitative - Real time PCR (qPCR)

The qPCR works on the same principle as RT-PCR but it measures the amplification of the gene of interest in real time. The value is expressed either in "Threshold Cycle" (C_T) or in "Crossing Point" (C_P) and describes the amplification cycle in the exponential phase where the fluorescence of the amplicon is higher than the background fluorescence. Different kinds

of reporter systems are available but here the SYBR Green dye was used, which can have the disadvantage of detecting all double strand DNAs including primer dimers. All samples were normalized against the housekeeper gene ß-Actin and water was used as a negative control. After the run was finished the results were quality checked by analysing the melting curves and C_P values. If the melting curve was not suitable or the C_P value was missing the measurement was repeated. If the C_P value of ß-Actin was ≥25, even after repeating the measurement, the sample was not taken into account for analysis due to poor RNA quality. If the melting curve was suitable, the C_P values of ß-Actin were right and the C_P values of the triplicates did not differ more than one value from each other the C_P values were taken for further analysis. To compare the expression between two genes first the median of the triplicates of ß-Actin and NY-BR-1 was determined. $-\Delta C_P$ (median ß-Actin–median NY-BR-1) was calculated. $2^{\wedge(-\Delta CP)}$ was calculated to receive the expression ratio between NY-BR-1 and ß-Actin (Schefe et al., 2006).

Table 3.9: Master mix for the qPCR/sample.

Components	Volume [µL]
2x SYBR Green Mix	5
„Forward"-Primer (10 µM)	0,2
„Reverse"-Primer (10 µM)	0,2
H_2O	2,6
Non - diluted cDNA	2

Table 3.10: Protocol to amplify specific sequences in qPCR. The best results for used ß-Actin and NY-BR-1 primer were obtained at the indicated annealing temperature. If other primers were used the annealing temperature was adjusted.

Step	Temperature	Time	Cycle
Initial denaturation	95 °C	120 s	
Denaturation	95 °C	15 s	40
Annealing	57 °C	25 s	
Elongation/Detection	72 °C	20 s	
Melting curve	62-95 °C		
Cooling	< 40 °C	Until remove	

3.2.5.15 Quantitative DNA Methylation Analysis by EpiTyper MassArray

Quantitative analysis of DNA methylation was performed using mass-spectrometry based EpiTyper MassArray technology (Sequenom, San Diego, USA) as described before (Ehrich et al., 2005). Using this system, methylation differences ≥5 % are detectable by quantifying median methylation of CpG units (single CpGs or two or more CpGs analysed together). 500

ng genomic DNA was subjected to sodium bisulfite treatment using the EZ DNA Methylation™Kit (Zymo Research, Orange, USA), according to the manufacturer's instructions. Subsequent PCR amplification and base-specific cleavage of nucleic acids enabled separation of methylated and unmethylated CpG units according to their mass to charge ratio (m/z). Genomic locations and positions of amplicons are summarized in Table x. The data were generated in the AG Gerhäuser, Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

*Table 3.11: Summary of MassArray amplicons of NY-BR-1. (*tss: transcription start site)*

Primer ID	Gene Name	Chr	Start	End	Distance to tss*	Left Primer	Right Primer	Length	CpGs
53	NYBR-1A	10	37413170	37413581	-1614	TTGGTTTATGTTTGTTTTGTAGTTTT	CCAAACTAATATATTCCCTTCTAATAAA	412	12
54	NYBR-1F	10	37413626	37413953	-1158	GATTTATGGGTTAGAAGTTTTTAGGG	AAATTCCAAAACTACCTAACCAACA	328	15
55	NYBR-1G	10	37413925	37414222	-859	TTTATGTTGGTTAGGTAGTTTTGG	CCCTAAAAACTTCTAACCCAAAAAT	298	14
56	NYBR-1I	10	37414198	37414505	-586	ATTTTTGGGTTAGAAGTTTTTAGGG	CAACTCAACAAACTAAAAAAACAC	308	21
57	NYBR-1L	10	37414771	37415126	-13	GTATTTTAAGTTTGGGGGTTTGTTT	ACCCTTTCAACCAACTAATCTATACC	356	12
58	NYBR-1M	10	37414556	37414797	-228	TGGTATAGATTAGTTGGTTGAAAGGG	AACCTAAATATACAAAAACTTAAC	242	19

3.2.5.16 Agarose gel electrophoresis

1 % (w/v) agarose gel (agarose + 1x TAE buffer (40 mM TrisHCl, pH 8.0, 20 mM acetate, 1 mM EDTA)) at 4-6 V/cm in a Sub-Cell-System (Biorad) was used to separate DNA fragments. To make the DNA fragments visible at 312 nm (UV light) 0.2 µg/mL ethidium bromide was administered to the liquid agarose. As a DNA marker a 100 bp or 1 kb ladder (New England biolabs, N3231L/N3232L) was used.

3.2.5.17 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

This technique is used to separate proteins according to their electrophoretic mobility. The mobility depends on the charge, length and conformation of the molecule.

6–12 % SDS–polyacrylamide gels were used depending on the size of the molecule of interest. The liquid gels were poured into 1 mm thick glass chambers (BioRad). The polymerized gels were placed into a Mini-PROTEAN-System (BioRad) filled with running

buffer. 5 x Laemmli buffer (Sigma, S3401-10VT) was added to the lysates. They were heated up for 5-10 min at 95 °C, cooled down on ice and shortly centrifuged. To separate the proteins 60 V was adjusted for the collecting gel and 130–145 V for the separation gel. The used ladder with a defined molecular weight was the peqGold Protein marker (peqLab, 27-2110).

Table 3.12: Composition of buffer used in SDS-Page.

Buffer	Composition
Separation gel buffer	1.5 M TrisHCl, pH 8.8
Collecting gel buffer	0.5 M TrisHCl, pH 6.8
1x SDS running buffer	25 mM TrisHCl, pH 8.3, 192 mM Glycin, SDS 0.1 % (v/v)

3.2.5.18 Western Blot

The "semi-dry" immunoblot system was used (Trans-Blot SD Chamber, BioRad) to transfer the proteins from SDS-PAGE to a Polyvinylidene fluoride (PVDF) membrane (Millipore, IPVH00010). Each blot contained blot paper (Chromatographypaper, GE, 3030-931), the PVDF membrane and the SDS gel. The transfer for NY-BR-1 lasted 2 h with a current of 70 mA/blot.

The general procedure consisted of blocking the PVDF membrane with blocking buffer for one hour at room temperature or overnight at 4 °C. The primary antibody was added diluted in blocking buffer and before adding the secondary antibody diluted in blocking buffer, the membrane was washed three times with PBS-Tween. The washing step was repeated after secondary antibody incubation to remove all unspecific antibody bindings. To receive a signal through an enzymatic reaction ECL plus (ECL Plus Western Blotting Detection System, GE healthcare, RPN2132) was added to the PVDF membrane and incubated for 1 min at room temperature. The ECL was removed, the PVDF membrane was placed into a cartridge, a film was put atop (Lucent Blue Film for Western Blot, Biozym) and incubated for 30 sec to 15 min. The film was developed in an AGFA CP1000 machine.

The polyclonal antibody α-NYBR-1 (C-18), and the monoclonal α-NY-BR-1#2 (mouse) antibody were tested on either cell lines (MCF-7) or primary material (stimulated samples, pleural effusions).

Table 3.13: Buffer composition used for Western blotting.

Buffer	Composition
Anode buffer I	300 mM Tris, Methanol 10 % (v/v)
Anode buffer II	30 mM Tris, Methanol 10 % (v/v)
Cathode buffer	30 mM Tris, 40 mM Glycin, Methanol 10 % (v/v)
PBS-Tween	1x PBS, Tween-20 (Polysorbate 20) 0.1 % (v/v)
Blocking buffer	1x PBS, Tween-20 (Polysorbate 20) 0.1 % (v/v), milk powder 5 % (w/v)

Table 3.14: Overview of primary/secondary antibodies used in Western blot with their dilutions.

Primary antibody	Dilution	Secondary antibody	Dilution
α-NYBR-1 (C-18)	1:200	Donkey α-goat HRP conjugated	1:5000
α-NY-BR-1#2 (Mouse)	1:500	Goat α-mouse IgG Pox conjugated	1:5000

3.2.5.19 Silver staining – Mass spectrometry compatible

After SDS–Page the gel was placed into a petri dish filled with fixation buffer (30 % ethanol, 10 % acetic acid, 60 % aqua dest.) and incubated overnight. The fixation buffer was removed and replaced with sensitize solution (0.3 % (w/v) potassium tetrathionate, 0.5 M potassium acetate, 30 % ethanol) for 45 min. The gel was washed six times 10 min with aqua dest. 0.2 % (w/v) silver nitrate solution was added and incubated for 1– 2 h. The gel was washed 15 sec by rinsed aqua dest. Developing solution (3 % (w/v) potassium carbonate, sodium thiosulfate-pentahydrate [10 %], formalin [37 %]) was administered and incubated between 5 – 10 min. The process was stopped by adding stop solution (4 % (w/v) tris, 2 % (v/v) acetic acid) for 45 min. At the end the gel was washed twice 30 min with aqua dest. The whole silver staining process was done at room temperature. A photo was taken and the gel was stored in aqua dest. until mass spectrometry analysis.

3.2.5.20 Protein complex immunoprecipitation (Co-IP)

To identify specific protein interaction partners of NY-BR-1, co-immune precipitations were performed on lysates of pc3-NY-BR-1 and pcDNA3-empty vector transiently transfected cells (HEK293, MCF-10A, and MCF-7). 24 h after transfection the cells were lysed as described before (protein analysis). 30 µl A/G Plus agarose beads (Santa Cruz, sc 2003) were combined with either 200 µg lysate (for pre-clearing) or with 2 µg α-NY-BR-1 (C-18) antibody. The samples were filled up to 500 µl final volume with Co-IP buffer plus 1 mM DTT and 1 % PIC. The beads were incubated for 2 h at 4 °C while rotating. All beads were centrifuged for 30 sec at 8000 rpm and 4 °C. The pre-cleared supernatant of the lysate incubation was filled

into a fresh pre-chilled tube. The beads were washed three times with 500 µl Co-IP buffer plus 1 mM DTT and 1 % PIC for 30 sec at 8000 rpm and 4 °C and stored as the first control at -80 °C. The supernatant of the samples, with the beads incubated with the antibody, was discarded and 450 µl of the pre-cleared lysate was added. The beads were incubated over night at 4 °C while rotating. The rest of the pre-cleared lysate was stored at -80 °C as a second control. 10 µl A/G Plus agarose beads were added additionally and incubated for 2 h at 4 °C while rotating. The beads were collected by centrifugation for 30 sec at 2500 rpm and 4 °C. The supernatant was stored as third control. The beads were washed three times with Co-IP buffer plus 1 mM DTT and 1 % PIC by inverting the tubes. The beads were centrifuged for 30 sec at 800 rpm, 4 °C and the supernatants were stored as fourth, fifth and sixth control. 5x Laemmli sample buffer was added to the beads and they were heated for 10 min at 95 °C. The beads were centrifuged for 30 sec at 2500 rpm, 4 °C and the supernatant was collected (IP reaction). The beads were washed three times with 500 µl Co-IP buffer plus 1 mM DTT and 1 % PIC for 30 sec at 8000 rpm, 4 °C and used as seventh control. Just the IP reaction of pc3-NY-BR-1 and pcDNA3-empty vector transiently transfected cells was applied in triplicates on a SDS-Page for mass spectrometry analysis. Western Blot and mass spectrometry compatible silver staining were used to detect proteins after electrophoretic separation on polyacrylamide gels. Finally, an analysis of the proteins was done by mass spectrometry.

The raw data, kindly provided from the Functional Proteome Analysis, German Cancer Research Center (DKFZ), INF 280, 69120 Heidelberg, Germany, contained information about accession number, protein description, score, mass [Da], protein matches, significant protein matches, protein sequences, significant protein sequences, and coverage [%]. The complete data set can be found in the laboratory of the Jäger group. To receive a better overview the raw data were analysed as followed: unique proteins in the NY-BR-1 fraction were taken and a cut-off was applied to a score of ≥ 40 and to significant protein sequences with a cut-off of at least two. Moreover, the unique proteins found in the NY-BR-1 fraction of each cell line were compared to each other to see which proteins are shared by either two or in all three cell lines.

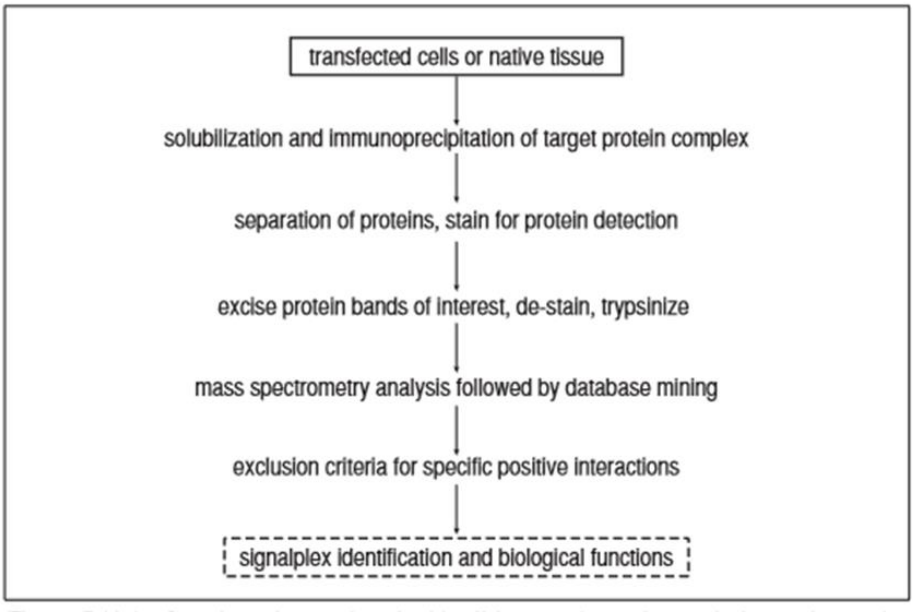


Figure 3.1: Workflow to identify specific interaction partner of a protein of interest (adapted from (Free et al., 2009)*).*

3.2.5.21 Immunohistochemistry (IHC) staining

Formalin-fixed paraffin embedded (FFPE) tissue was used for the serial tissue sections and cryo-preserving for the stimulated tissue pieces. Therefore, the FFPE tissue, collected from breast reductions, was fixed overnight in 4 % buffered formalin and then placed into a tissue processor (Microm, STP 420D) (protocol see appendix). After sample processing the FFPE block was stored at room temperature until cutting. The thickness of the tissue section was between four-six µM.

The tissue was fixed with Tissue-Tek OTC-Mounting medium (Sakura, Staufen) for tissue sections of cryo-preserved tissue. The tissue sections with 8-12 µm thickness were produced in a Kryostat-Mikrotom Microm 550 MVP (Thermo Fisher Scientific, Dreieich) between -20 - -24 °C depending on the fat and connective tissue amount in the tissue itself. The tissue sections were brought to HistoBond+ (Marienfeld, Lauda Königshofen) slides. They were stored at -80 °C until staining. For staining the sections were fixed with 4 % PFA 20 min at room temperature in the dark. Afterward the slides were washed 6 times 10 min with 1x PBS. MeOH/EtOH was added for 10 min in a 1:3 ratio and the slides were dried at room temperature. At last TTX was added for 5 min.

The staining of cryo-preserved or FFPE tissue with different kind of antibodies was done with the BOND-Max (Leica, Nussloch) detection system. The stained slides were scanned with NanoZoomer2.0-HT (Hamamatsu Photonics, Herrsching) and viewed with NanoZoomer Digital Pathology Virtual Slide Viewer-Program (Version 1.2.36, Hamamatsu Photonics, Herrsching).

Table 3.15: Antibodies with their specific dilutions and retrievals used for IHC staining on FFPE tissue. ER1 (Leica, AR9961) citrate based buffer, ER2 (Leica, AR9640) an EDTA based buffer.

Antibody	Dilution	Retrieval
ER α	1:1000	ER1
Progesterone	1:250	ER1
Ki-67	1:100	ER2
α-NY-BR-1#2	1:50 (stock: 0.1 µg/µl)	ER1

3.2.6 *In silico* analysis

3.2.6.1 Transcription factor binding

By using the tool Alibaba 2.0, transcription factor binding sites can be predicted. The Alibaba algorithm uses a dataset of known binding sites to predict potential binding sites in any input sequence. This approach was shown to achieve high accuracy (Grabe, 2002). The Alibaba 2.0 algorithm was used to predict transcription factor binding sites in the ANKRD30A promoter up to 2000 bp upstream of transcription start site.

3.2.6.2 SNP Mining

dbSNP is hosted by the National Center for Biotechnology Information (NCBI) and is the largest repository of SNP data with over 140 million submitted variations.

Another source of variation data is provided by the "The National Heart, Lung and Blood Institute" (NHLBI). The protein coding regions of each individual genome (i.e. exome) is sequenced and the variation data is made publicly available (Dorschner et al., 2013).

The Ensembl Variation database incorporates variation data from several sources including dbSNP and NHLBI ESP. The web interface MartWizard (http://central.biomart.org/martwizard/) of the BioMart Central Portal was used.

The Ensembl transcript ID ENST00000611781 of the ANKRD30A gene was used to retrieve all available germline variations together with the corresponding genomic coordinates, the variant descriptions, and the validation status. Using the variant descriptions, coding non-synonymous SNPs (nsSNPs), coding synonymous SNPs (sSNPs) and intronic SNPs were filtered.

3.2.6.3 Prediction of the functional impact of coding nsSNPs using SIFT, PROVEAN, and Polyphen

The tools SIFT and PROVEAN are available online at http://provean.jcvi.org/. On the website, the tool PROVEAN Human Genome Variants, which provides PROVEAN and SIFT predictions for a list of human genome variants were used. The list of genomic coordinates and variants of our filtered 191 nsSNPs were submitted, and the default threshold of delta score <= -2.5 to detect deleterious variations was chosen. PolyPhen-2 is available online at http://genetics.bwh.harvard.edu/pph2/. Here, the option 'Batch query' was taken and the list of genomic coordinates and variants of our filtered 191 nsSNPs was submitted.

The prediction tool SIFT evaluates the functional impact of SNPs based on sequence homology. The prediction is based on the degree of conservation of each amino acid residue of the query sequence. To assess the degree of conservation, SIFT compiles a dataset of functionally related protein sequences by searching the protein databases UniProt and TrEMBL using the PSI-BLAST algorithm and builds an alignment of the found sequences and the query sequence. In the second step a normalized probability for each substitution at each position of the alignment is calculated and is then recorded in a scaled probability matrix. This scaled probability is also called the SIFT score and a substitution is considered to be tolerated if the score is greater than 0.05; those less than 0.05 are predicted to be deleterious. The SIFT approach assumes that a highly conserved position is intolerant to most substitutions, whereas a poorly conserved position can tolerate most substitutions.

The tool PROVEAN also uses an alignment approach to assess the functional impact of SNPs. For this, a set of homologous and distantly related sequences from the NCBI NR protein database is collected using BLASTP. To remove redundancy, the collected sequences are clustered, based on a sequence identity of 80 %. A so called supporting set of sequences is assembled by adding sequences from clusters most similar to the query sequence, until a sufficient number of clusters are achieved in the supporting set. In the second step, for each sequence in the supporting sequence set, a delta score is computed using the BLOSUM62 substitution matrix. For each cluster, an average delta score is computed, and the averaged delta scores are again averaged among all clusters. This unbiased averaged delta score is the final PROVEAN score. The impact of a variation on protein function can be measured as the change in alignment score, the delta score. Low delta scores are interpreted as variations leading to a deleterious effect on protein function, while high delta scores are interpreted as variations with neutral effect.

PolyPhen-2 combines information on sequence features, multiple alignments with homologous proteins, and structural parameters to predict the impact of a SNP on protein function. For sequence-based assessment, PolyPhen-2 tries to identify the query as an entry in the UniProtKB/Swiss-Prot database. Using the feature table of the corresponding entry, PolyPhen-2 checks if a given SNP occurs at functional relevant site, e.g. if the SNP lies within a transmembrane, signal peptide, or binding region. For each variant PolyPhen-2 calculates a position-specific independent counts (PSIC) score. The PSIC score difference between the two variants describes the impact of a particular amino acid substitution: the higher the PSIC score difference, the higher functional impact the substitution is likely to have. A BLAST query of the query sequence against protein structure databases is carried out to identify corresponding 3D protein structures. If corresponding structures are found, they are used to assess, whether the SNP is likely to destroy the hydrophobic core, interactions with ligands or other important features of the protein. Finally, all parameters are taken together and empirical prediction rules are applied to make the final decision, whether the SNP is damaging or benign.

3.2.6.4 NY-BR-1 expression in breast cancer subtypes (Breast Cancer RNA-Seq Dataset)

Varley *et al.* published a RNA-Seq study on recurrent read-through fusion transcripts in breast cancer. 168 samples were analysed in this study, including samples from estrogen positive (ER^+) and triple negative breast cancer (TNBC), as well as samples from uninvolved adjacent tissues (Varley et al., 2014). All data generated in this study is available for download from the NCBI Gene Expression Omnibus (GEO) through accession number GSE58135. RNA-Seq raw reads were downloaded from the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) and aligned against the Gencode reference genome using STAR aligner. Subsequently, RNA abundances were assessed using the transcript per million (TPM) calculation. Significance of NY-BR-1 expression differences between the sample groups was assessed using a two-sided Student's t-Test (Varley et al., 2014).

4 Results

The breast cancer associated antigen NY-BR-1 fulfils several criteria to be a target for immunotherapy. NY-BR-1 over-expression in primary breast tumours as well as in metastases and spontaneous humoral responses has been found. But up to date the biological function in healthy and breast tumour tissue is unknown. Therefore, our study aimed at describing this protein in more detail. The results of the molecular and functional characterization of NY-BR-1 are presented in this chapter.

One main focus in our study was to analyze how NY-BR-1 is transcriptionally regulated by combining *in silico* data with wet lab techniques. Functional assays like proliferation assay and cell cycle analysis were performed to observe the effect of NY-BR-1 overexpression in four transiently transfected cell lines. Co-IP of transiently transfected cells lines was done, followed by mass spectrometry analysis to get evidence for potential interaction partners of NY-BR-1. IHC stained healthy breast tissue and mammosphere generation was done to look at NY-BR-1 expression in potential progenitor cells. *In silico* analysis was administered to receive an overview of SNPs in the ANKRD30A gene and its assumed promoter, to show the exon coverage and to correlate NY-BR-1 expression data with methylation status, and expression data of other genes. An overview of the performed experiments is depicted in figure 4.1.

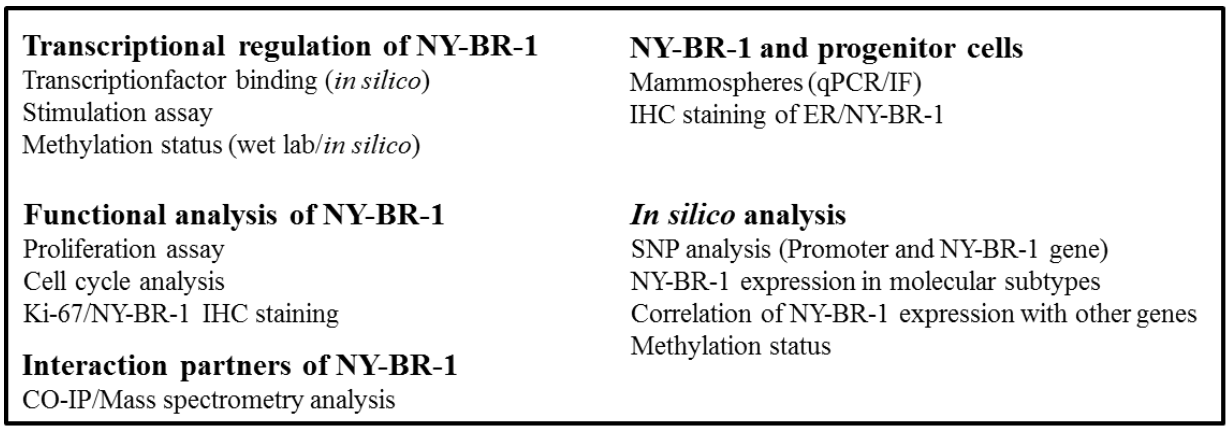


Figure 4.1: Overview of the experiments to characterize NY-BR-1 on molecular and functional levels.

4.1 NY-BR-1 expression pattern

The NY-BR-1 expression pattern was analysed before on mRNA level in a panel containing different normal tissues. The results showed that NY-BR-1 is strongly expressed in the

mammary gland and testis (Jäger et al., 2001). Several different studies also looked at the NY-BR-1 expression pattern in breast carcinoma samples. In one of the first studies 25 breast cancer samples were analysed with 21 samples being positive for NY-BR-1 on mRNA level (Jäger et al., 2001). Using quantitative and qualitative RT-PCR the second study could demonstrate that NY-BR-1 is over-expressed in 80 % of the patients. An additional microanalysis revealed that NY-BR-1 was up-regulated >2-fold in 69 % of the patients. IHC stainin in normal breast tissue revealed that the ductal epithelium of the mammary gland is weakly positive for NY-BR-1 (Seil et al., 2007).

The endogenous NY-BR-1 expression was analysed in twelve different breast cell lines by qRT-PCR and Western blot. Four cell lines (MDA-MB-453/HTB 131, BT474, T47D, ZR-75-30) were positive for NY-BR-1 expression on RNA level but this finding could not be confirmed on protein level (Seil, 2005). The publicly available RNA-Seq dataset by Varley *et al.* contained the RNA expression profile of 28 samples of breast cancer cell lines (Varley et al., 2014). Only in six breast cell lines (BT-474, MDA-MB-134, DY30T2, MDA-MB-361, ZR-75-30, MDA-MB-453) a weak NY-BR-1 gene expression was detected on RNA level.

MCF-10A, MCF-7, HEK293 and HEK293T cell lines, used for the subsequent experiments, were tested in either qPCR or western blot. Except for MCF-7 (weakly NY-BR-1^{+}) all cell lines were negative on RNA level (figure 4.2). NY-BR-1 could be detected on protein level in MCF-7 cells. Thus, no suitable cell culture system was available expressing endogenous NY-BR-1. Therefore, the cells needed to be transiently transfected to peform certain experiments with facing the challenge that the transfection efficacy varies strongly between the cell lines and NY-BR-1 expression decreases rapidly after three to four days (figure 4.23).

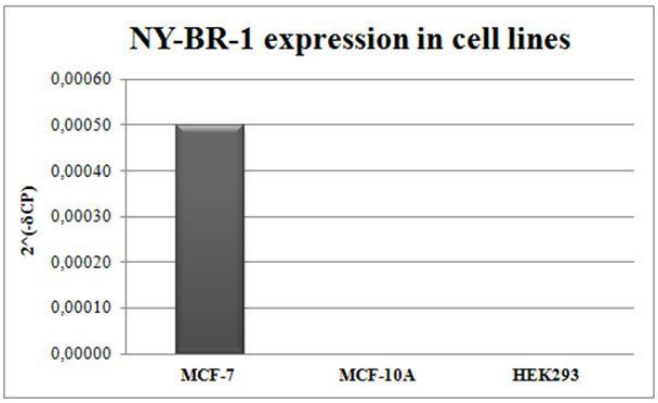


Figure 4.2: NY-BR-1 expression was analyzed by qPCR in MCF-7, MCF-10A and HEK293 cells. The NY-BR-1 expression was normalized against the housekeeping gene ß-Actin. HEK293T cells were tested before by Inka Seil (Seil, 2005).

Not only was the cell culture system a limiting factor but also the available tools for NY-BR-1 detection in e.g. Western Blot. Beside one mon(Jäger et al., 2007) a polyclonal antibody (α-NY-BR-1 (C-18)) is commercially receivable. Unfortunately, this antibody shows unspecific bindings in Western Blot. Also with α-NY-BR-1#2 antibody it was challenging to detect endogenous NY-BR-1 in Western Blot in different settings, e.g. stimulated healthy samples (figure 4.3A) and pleural effusion cells (figure 4.3B). It is noteworthy to mention that the NY-BR-1 band of transfected cells is approximately at 180 kDa (A: lane 2, B: lane 1), whereas endogenous NY-BR-1 can be found at approximately 160 kDa (B: lane 3).

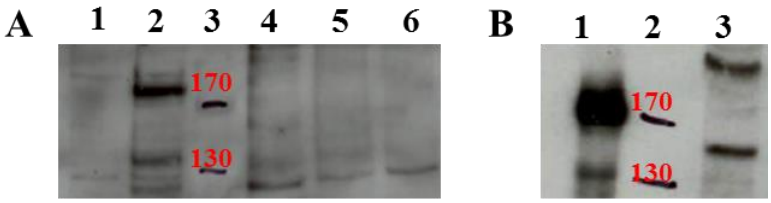


Figure 4.3: Endogenous NY-BR-1 expression was analyzed by Western Blot in different samples. The α-NY-BR-1#2 antibody was used (A) 1:HEK293 transfected with pc3 empty vector, 2: HEK293 transfected with pc3 NY-BR-1, 3: Ladder [kDa], 4: HD-T-203 not cultured, 5: HD-T-203 TSA, 6: HD-T-203: Aza/VitD3, (B) 1: HEK293 transfected with pc3 NY-BR-1, 2: Ladder [kDa], 3: HD-A-90 Progesterone

The α-NY-BR-1#2 antibody is suitable for IHC staining and FACS analysis as shown in figure 4.4/figure 4.5 and as Seil *et al.* already reported (Jäger et al., 2007; Seil et al., 2007).

4.1.1 IHC staining of healthy breast tissue

Normal breast tissue is abounding in fat and connective tissue (figure 1.1). This can be a problem for cryo-preserved tissues: the production of tissue sections is difficult because the tissue section coils up while cutting and in the automated staining process (BOND system, Leica) the tissue sections tend to get off the slides or the tissue sections look disruptive. In normal breast tissue, only few luminal epithelial cells express NY-BR-1 as assessed by immunohistochemical staining with the α-NY-BR-1#2 antibody on FFPE tissue. In figure 4.4 stained FFPE tissue sections of nine different healthy patients (HD-T-203 - HD-T-316) are shown. As indicated with the arrows the NY-BR-1 expression pattern is mosaic like.

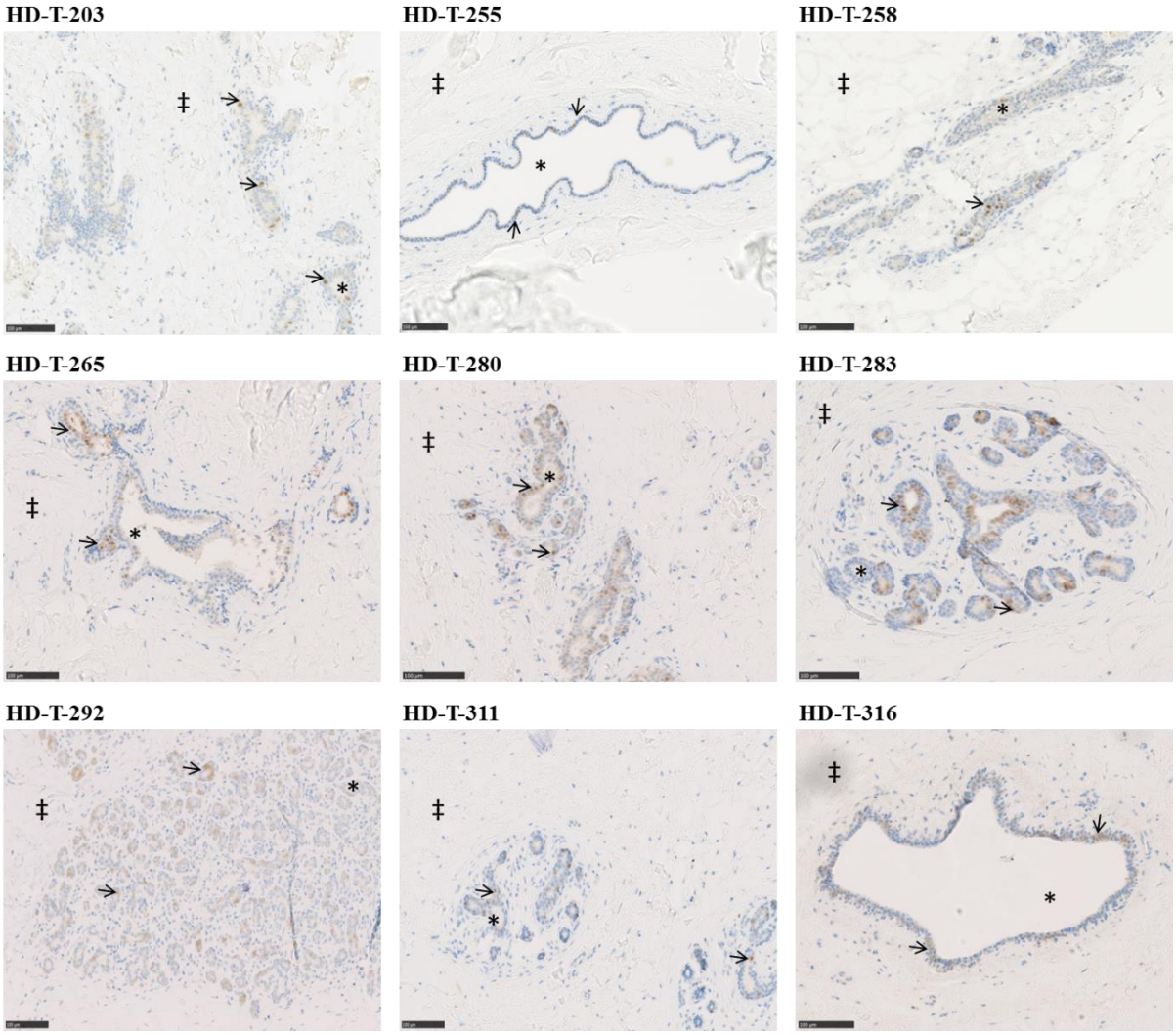


*Figure 4.4: IHC staining of FFPE tissue sections derived from nine healthy patientss. The nucleus is stained blue (haematoxylin) and the brown stained cells are NY-BR-1 positive. The tissues derived from breast reductions. The magnification is 20 fold. (‡: connective tissue, *: lumen of mammary gland, arrow: NY-BR-1 positive cells)*

4.1.2 IHC staining of breast cancer tissue or metastases

The staining of breast cancer samples and metastases show a different NY-BR-1 expression pattern than in healthy breast tissue. Based on the fact that the different progenitor subsets in the mammary gland might be responsible for giving rise to the different molecular subtypes of breast cancer the expression level of NY-BR-1 will depend on the molecular subtypes. Therefore, some patients are negative or weakly positive for NY-BR-1 (1G01UQ, 6617), whereas patients HD-T-202 and 6646 are highly positive for NY-BR-1 expression. Patients 6311 and 1OH7FL show a heterogeneous NY-BR-1 expression (figure 4.5).

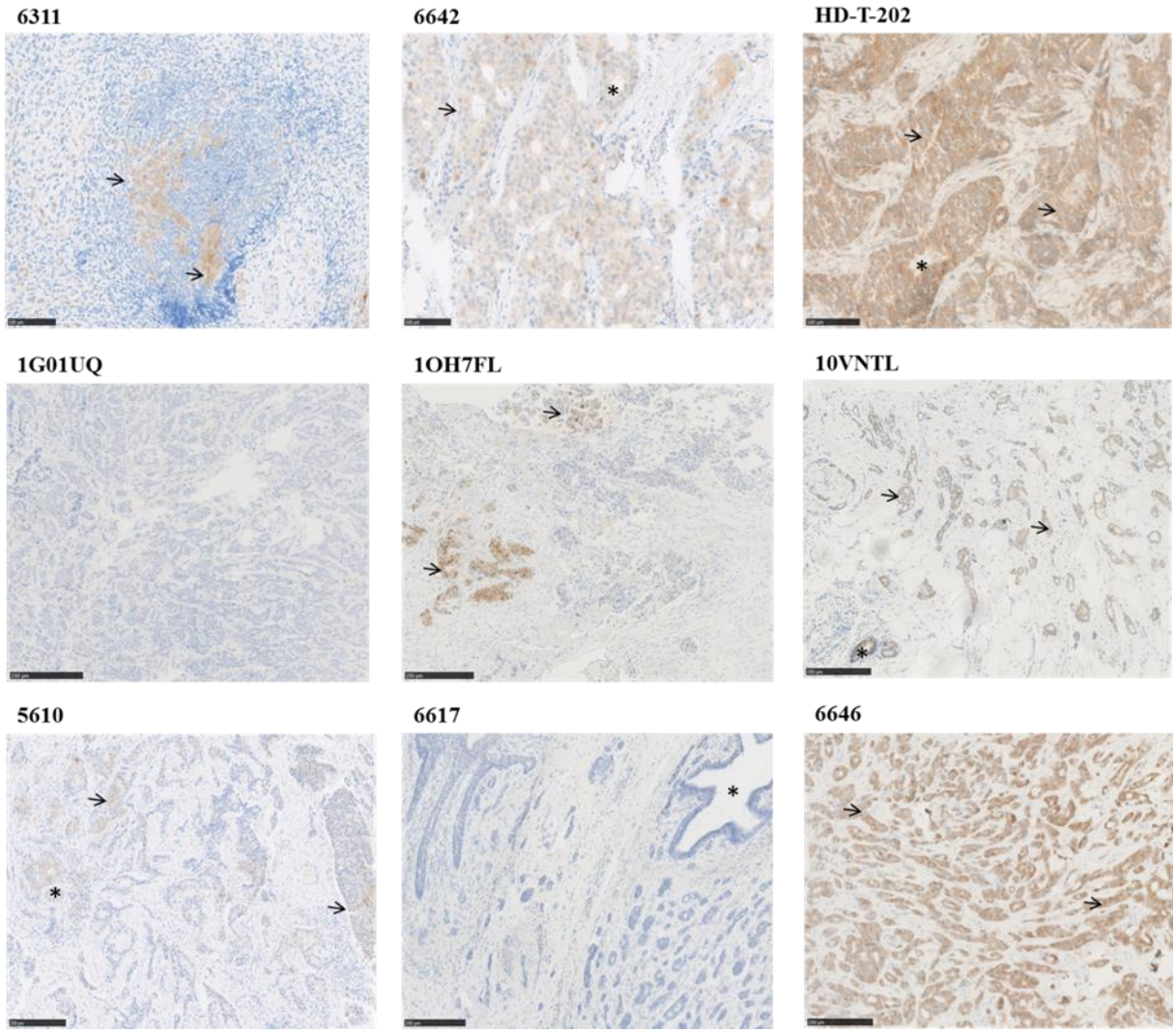


Figure 4.5: IHC staining of FFPE tumour samples or metastases (HD-T-202) of breast carcinoma patients. The nucleus is stained blue (haematoxylin) and the brown stained cells are NY-BR-1 positive. The magnification is 10 or 20 fold.) (: lumen of potential mammary gland, arrow: NY-BR-1 positive cells)*

4.2 How is NY-BR-1 transcriptionally regulated?

4.2.1 *In silico* analysis of transcription factor binding

Transcriptional regulation is a process occurring on many different levels. If a gene is transcribed depends on the methylation status of histones and CpG islands, transcription factor binding, SNPs and many more factors. The use of the Alibaba 2.0 tool formed the basis to get a more detailed look in the transcriptional regulation of the NY-BR-1 gene. This tool predicts transcription factor binding sites in the assumed promoter region of ANKRD30A as well as binding sites for nuclear hormone receptors and CpG islands (figure 4.6) (Grabe, 2002). Table 4.1 summarizes all predicted transcription factors with the respective number of

binding sites in the NY-BR-1 promoter region up to 2000 bp upstream of the transcription start site.

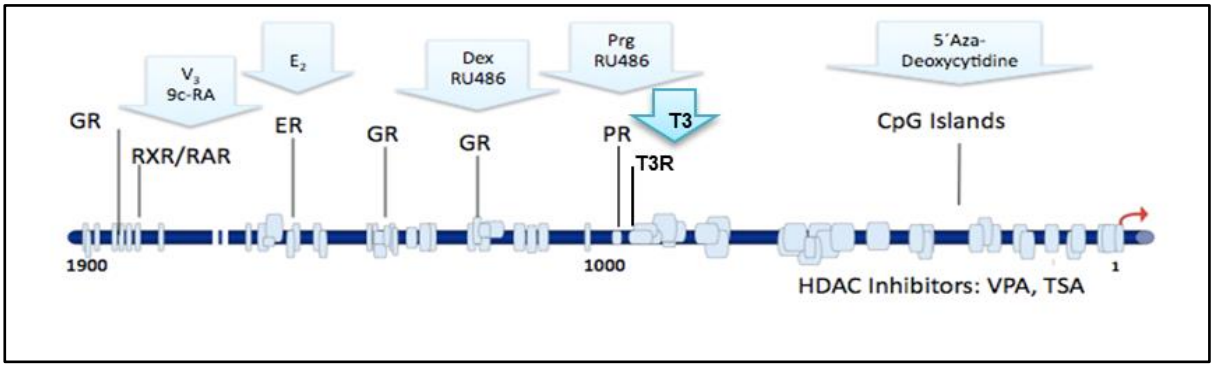


Figure 4.6: Predicted binding sites for nuclear hormone receptors and CpG islands in the promoter region of NY-BR-1 with Alibaba 2.0. (CpG Islands: genomic regions that contain a high frequency of CpG sites; HDAC Inhibitors: histone deacetylase inhibitors; VPA: valproic acid; TSA: trichostatin A; and T3R: thyroid hormone receptor)

Table 4.1: Overview of all predicted transcription factors with the number of binding sites in the ANKRD30A promoter region by using the Alibaba 2.0 algorithm.

Transcription factor	Binding sites	Transcription factor	Binding sites	Transcription factor	Binding sites	Transcription factor	Binding sites
Adf-1	2	E4	2	HOXA4	1	Oct-1	1
AP-1	5	E47	1	HSF	2	Oct-2.1	1
AP-2	2	Egr-1	3	ICSBP	1	Pit-1a	1
AP-2α	7	ER	1	Krox-20	5	PR	1
ATF	3	ETF	1	Max1	2	RAP1	4
C/EBPα	13	GATA-1	3	MEB-1	2	RARβ	1
C/EBPβ	3	GBF1	1	MEF-2	1	represso	1
C/EBPγ	1	GCN4	1	Myf-3	1	RXRα	1
c-Fos	3	GR	3	MyoD	4	SP1	68
c-Jun	4	Hb	5	myogenin	2	T3R	1
c-Myc	3	HEB	1	NF-1	18	TBP	2
c-Rel	1	HNF-1	2	NF-muE1	2	USF	3
dioxin r	1	HNF-1C	1	NF-κ	1	WT1	1
Dl	1	HNF-3	5	NF-κB	2	YY1	4
E1	4	HNF-3B	1	NRF-1	1		

Besides the nuclear hormone receptor binding sites, other interesting transcription factors such as AP-1, C/EBPα, C/EBPβ, C/EBPγ, c-Fos, c-Jun, c-Myc, GATA-1, HNF-3, Max1, SP1 and YY1 have predicted binding sites in the promoter region. But the main focus in this project was on the nuclear hormone receptor binding sites because mammary glands are under

hormonal influence. Furthermore, Theurillat *et al*. could show that NY-BR-1 is co-expressed with ER in breast tumours (Theurillat et al., 2007).

4.2.2 Stimulation assay

To study the hormonal influence on transcriptional activation *in vitro*, several NY-BR-1 negative or weakly positive breast cancer cell lines (T47-D, 226L, ZR-75-1, MCF-7) were stimulated for 72 h with different hormones and other agents (demethylation agent, histone deacetylation inhibitors) either alone or in combination. In these preliminary data NY-BR-1 expression was analyzed by quantitative PCR with NY-BR-1 copy number normalization against the housekeeper gene cyclophilin-B. In figure 4.7, results of MCF-7 (weak NY-BR-1^{+}) and ZR-75-1 (NY-BR-1^{-}) are shown as an exemplary representation of our data, generated by Thomas Giese (Department of Orthopaedics and Institutes of Immunology and Pathology, University of Heidelberg, Germany). The sequences of the used primer are described in Wallwiener *et al* (Wallwiener et al., 2011). Compared to the copy number of cyclophilin-B (~1000) the NY-BR-1 copy numbers range from one to 63, which is very low. This suggests that by stimulation with the respective agents no considerable transcriptional activation of NY-BR-1 was achieved.

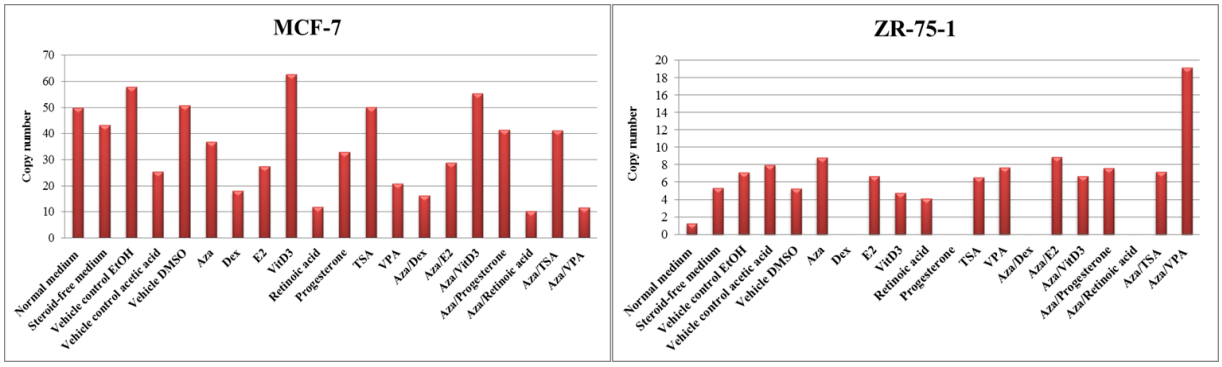


Figure 4.7: NY-BR-1 expression was analyzed in MCF-7 and ZR-75-1 cells after 72 h of incubation with 1µM 5'Aza, 200nM TSA, 1mM VPA and different hormones (10 nM E2, 1 µM progesterone, 1 µM Dex, 1 nM VitD3, 1 µM retinoic acid). Cells stimulated just with the vehicle are serving as controls.

The next step was to repeat the stimulation assays with normal mammary gland tissues derived from breast reductions. The idea was that in contrast to cell lines, primary tissues reflect the *in vivo* situation and maintain the natural environment of breast epithelial cells. Therefore, fresh needle biopsies of healthy mammary gland tissue derived from breast reductions of different female patients (18 to 64 years) were used. The tissue pieces were

cultured in phenol-free DMEM/F12 medium fully covered and stimulated over 2 to 48 hours with either 5'Aza, Dex, E2, VitD3 or in combination. The samples were then cryo-preserved and tissue sections were prepared for either immunohistological staining or RNA isolation to perform a quantitative PCR. Non-stimulated cultured tissue pieces as well as snap-frozen tissue pieces served as controls. As described at the beginning the cryo-preserved tissue sections were difficult to bring on a slide and were disrupted after staining. Therefore, a detailed analysis on protein level was not possible.

For the qPCR analysis the following primer were used for ß-Actin and NY-BR-1 because the same annealing temperature could be used (the binding sites of the NY-BR-1 primer within the gene can be found in the appendix (figure 6.1)):

Table 4.2: Oligos used for expression analysis with filemaker number of the AG Jäger laboratory; name of the gene of interest; forward/reverse sequence, melting temperature and size of the product.

Filemaker number	Gene	Oligo-Sequence	TM [°C]	Size[bp]
67 68	ß-Actin	FW: 5´-GCGGGAAATCGTGCGTGACATT-3´ RV: 5´-GATGGAGTTGAAGGTAGTTTCGTG-3´	56.7 55.7	232
100 80	NY-BR-1	FW: 5'-GAGGTGCTCCATCAACCACTTTC-3' RV: 5'-TCATGAGTTTTCTGTTTCTGCTTTC-3'	57.1 52.8	456

Other NY-BR-1 primer combinations were tested as well but they were excluded from analysis for three reasons:

1. The water control was positive because of primer dimers and SYBR Green detects all double strands.
2. The annealing temperature of both primer sets (NY-BR-1, housekeeping gene) was not compatible with each other.
3. The melting curves of other primer combinations approved.

4.2.2.1 Stiumulation of tissue pieces

The stimulated tissue pieces were analysed via qPCR and IHC staining to overcome the heterogeneity of the tissue and to detect a difference of NY-BR-1 expression on protein level. The cryo-preserved tissue pieces could not be IHC stained in an adequate way because some of the tissue sections did not stick on the slides while staining was performed and due to high amount of fat and connective tissue the tissue sections were disrupted. Therefore, no cyrp-preserved stimulated tissue pieces are shown.

First, the two control tissues in culture without any additives (w.a.) versus snap-frozen tissues of tissue pieces derived from breast reductions of healthy patients are compared regarding

their NY-BR-1 expression (figure 4.8). The NY-BR-1 expression was normalized against the housekeeping gene ß-Actin because it is not regulated by NY-BR-1. The formula for normalization is described in 3.2.5. The C_P values of ß-Actin range in the tissue pieces between 16 and 24, in the isolated epithelial cells between 15 and 23, and in the pleural effusion cells between 14 and 24.

The results indicate that NY-BR-1 expression is lost due to culturing because the non-cultured tissues have in almost all cases (except patient HD-T-177) a higher NY-BR-1 expression (C_P: 24-28) compared to the cultured tissues (C_P: 26-31). It is of note that the tissues are very heterogeneous and differ in the amount of glands and therefore also in the amount of epithelial cells, which are considered to be involved in NY-BR-1 expression. To compare the effects on NY-BR-1 expression after applying certain agents the cultured but not stimulated tissues (w.a.) were taken as a control.

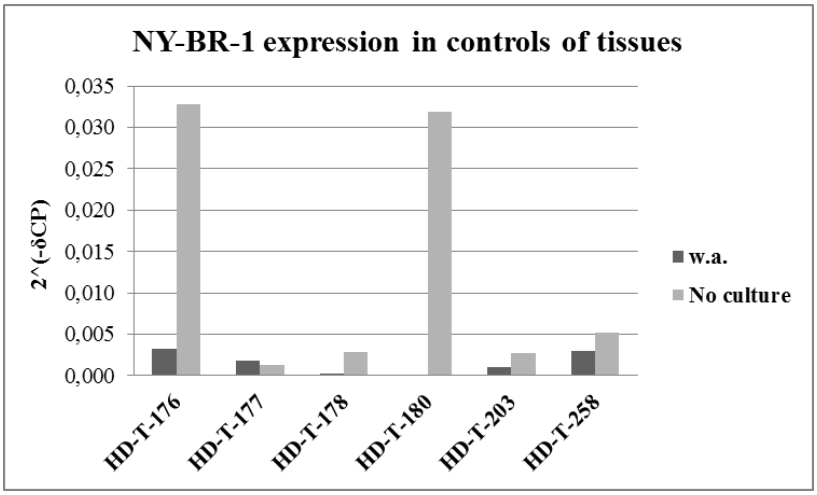


Figure 4.8: NY-BR-1 expression in the controls of tissue pieces derived from healthy patients after breast reductions. The dark grey bars (without additives (w.a.)) represent the NY-BR-1 expression in tissues cultured as long as all tissues but no supplements were given. The light grey bars show the NY-BR-1 expression in tissues never being in culture.

The results of the hormonal stimulation of the tissue pieces are represented in figure 4.9A. It seems that 5'Aza (C_P: 24) can up-regulate the NY-BR-1 expression in patient HD-T-176 compared to the non-stimulated but cultured cells (w.a.). In all other patients no up-regulation in NY-BR-1 expression can be detected by adding 5'Aza (C_P: 27-31). E2 stimulation in patient HD-T-177 (C_P: 29) can cause a NY-BR-1 up-regulation but in all other patients (HD-T-176, -178, -180, -258) (C_P: 27-32) this effect could not be observed. Stimulation with Dex (C_P: 27) has in patient HD-T-176 no effect on NY-BR-1 expression on RNA level. In all other

patients Dex and VitD3 down-regulate NY-BR-1 expression (HD-T-176 – HD-T-203) (Dex: (C_P: 27-32), VitD3 (C_P: 28-35)).

Dex, E2, and VitD3 were also given in combination with the demethylation agent 5'Aza (figure 4.9 B). In two out of three patients the combination Aza/Dex (C_P: 34/27) and Aza/E2 (C_P: 32/33) has a down regulatory effect on NY-BR-1 expression compared to the control. The level of NY-BR-1 expression increases as a result of Aza/VitD3 stimulation in two patients (HD-T-178/HD-T-180) (C_P: 31/26) but the other two (HD-T-203/HD-T-258) (C_P: 28/33) show a decrease compared to the control. These opposed effects can be explained by the heterogeneity of each tissue piece and the amount of epithelial cells varying from patient to patient.

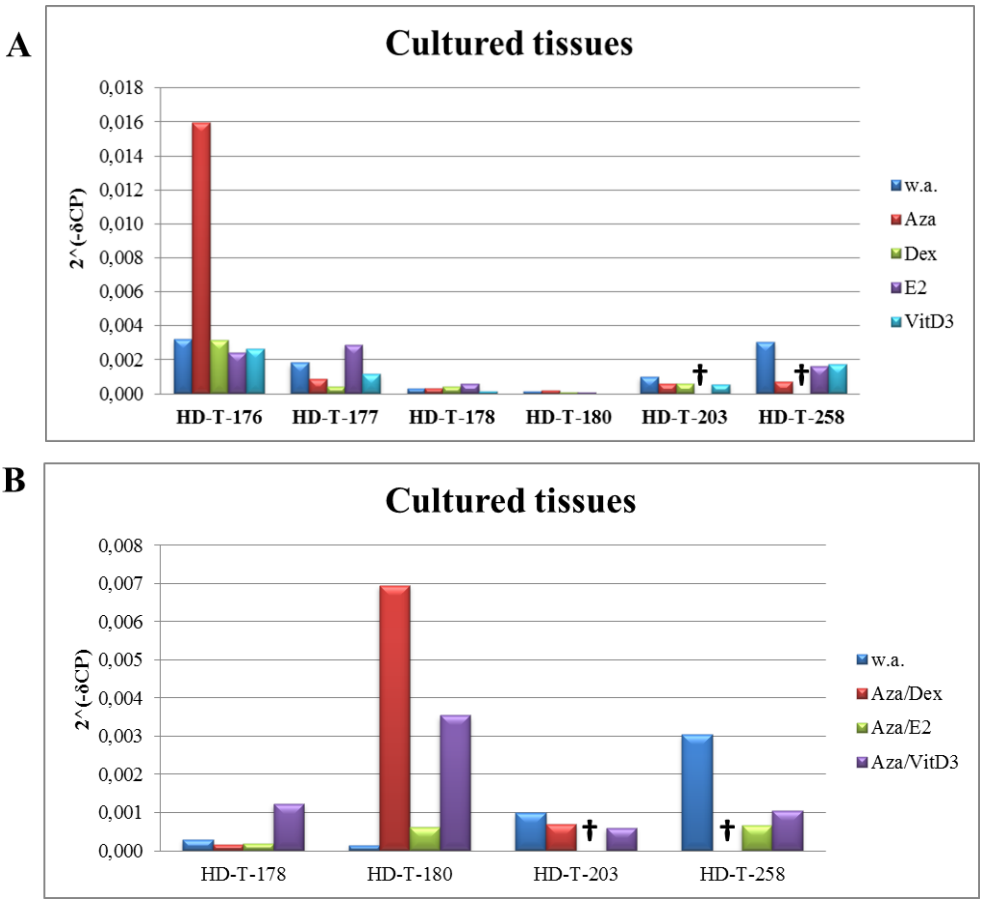


Figure 4.9: A: Up- or down-regulation of NY-BR-1 expression of cultured tissue pieces derived from breast reductions of six healthy patients by adding specific substances. The non-stimulated tissue piece functions as control (w.a.) B: The stimulation of cultured tissue pieces with hormones combined with 5'Aza. (1µM 5'Aza, 1 µM Dex, 10 nM E2, 1 nM VitD3) (†: no value available)

Depending on the amount of obtained material, additional stimulations were carried out with either 1 µM progesterone, 1 µM retinoic acid, 200 nM TSA or in combination with 5'Aza (figure 4.10). The single agents down-regulate the NY-BR-1 expression in both patients,

whereas the combinations Aza/Progesterone (C_P: 27/28) and Aza/TSA (C_P: 25/31) up-regulate NY-BR-1. The combination Aza/Retinoic acid in patient HD-T-203 (C_P: 27) results in a weak up-regulation, in patient HD-T-258 (C_P: 31) in a down-regulation of NY-BR-1 compared to the control.

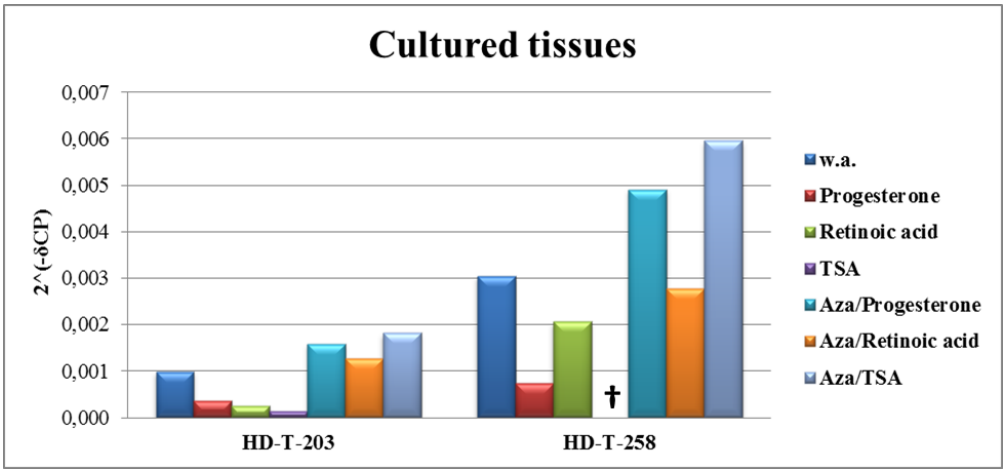


Figure 4.10: Results of the stimulation experiments of cultured tissue pieces derived from breast reductions of two healthy patients with additional substances. The non-stimulated tissue piece functions as control (w.a.) (1 µM 5'Aza, 200 nM TSA, 1 µM progesterone, 1 µM retinoic acid (†: no value available).

4.2.2.2 Stimulation of isolated epithelial cells

As mentioned before the tissue pieces are very heterogeneous and the amount of glands can vary between each patient and even between the single tissue pieces. Moreover, during preparation it is difficult to distinguish between connective and glandular tissue. To receive a more homogeneous cell population it was decided to isolate epithelial cells from tissues derived from breast reductions of healthy patients. With this step connective tissue and fat cells will be diminished after adding a digestion mix containing collagenase (digestion of collagen in the ECM and hyaluronidase (digests hyaluronic acid, an important component of ECM of connective tissues.

The single epithelial cells (luminal and myoepithelial cells) were taken into culture and stimulated with hormones or other agents for four to five days soon after isolation. Two patients could be analysed so far.

As described before the two control cells in culture without any additives (w.a.) versus snap-frozen cells are compared regarding their NY-BR-1 expression (figure 4.11). The results indicate that NY-BR-1 expression is lost due to culturing because in both cases the non-cultured cells have a higher NY-BR-1 expression (C_P: 26/25) compared to the cultured tissues (C_P: 26/35).

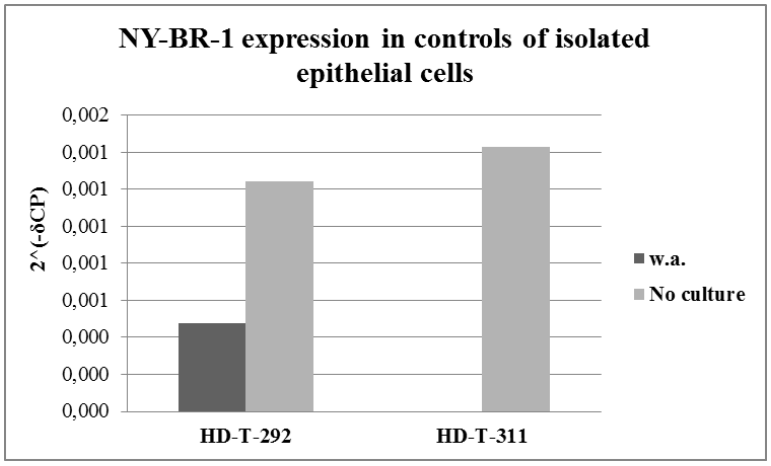


Figure 4.11: NY-BR-1 expression in the controls of isolated epithelial cells derived from healthy patients after breast reductions. The dark grey bars (without additives (w.a.)) represent the NY-BR-1 expression in cells cultured as long as all/cells but no supplements were given. The light grey bars show the NY-BR-1 expression in cells never being in culture.

After stimulation with Aza (C_P: 27), E2 (C_P: 27) and Tamoxifen (C_P: 35) a decrease in NY-BR-1 expression can be observed in patient HD-T-292. With all other substances (Dex (C_P: 26), VitD3 (C_P: 26), progesterone (C_P: 25), retinoic acid (C_P: 26), TSA (C_P: 26), VPA (C_P: 28)) an up-regulation is detected compared to the control cells. The control cells of patient HD-T-311 are NY-BR-1$^-$ (C_P: 35). In patient HD-T-311 an up-regulation of NY-BR-1 can be observed with all applied agents with the highest increase induced by Tamoxifen (C_P: 30) and the lowest by retinoic acid (C_P: 31) (figure 4.12A). In the combination setting with 5'Aza almost no change is visible for patient HD-T-311 compared to the single stimulation except for Aza/Tamoxifen (C_P: 35), which results in no detectable NY-BR-1 expression on RNA level. The same effect can be observed in patient HD-T-292. Aza/E2 (C_P: 31) and Aza/retinoic acid (C_P: 29) lower NY-BR-1 expression whereas all other combinations increase it compared to the non-stimulated cells in patient HD-T-292 (figure 4.12B). It is noteworthy that the general NY-BR-1 expression in both patients is very low due to a mixed cell population (myoepithelial/luminal cells) and not all cells express NY-BR-1 (figure 4.4).

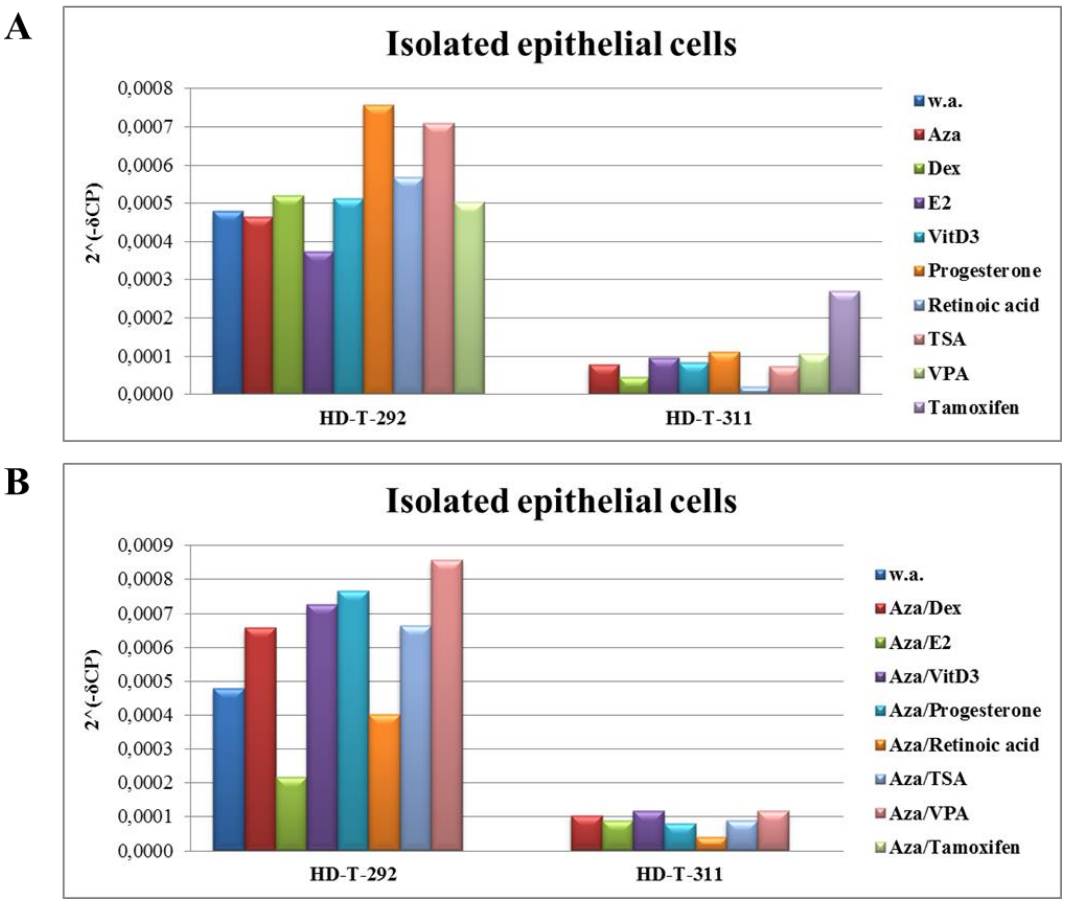


Figure 4.12: A: The effect on NY-BR-1 expression of isolated epithelial cells derived from breast reductions of two healthy patients by adding specific substances. The non-stimulated isolated epithelial cells functions as control (w.a.) B: The stimulation of isolated epithelial cells with hormones combined with 5'Aza. (1 µM 5'Aza, 1 µM Dex, 10 nM E2, 1 nM VitD3, 200 nM TSA, 1 mM VPA, 1 µM progesterone, 1 µM Dex, 1 nM VitD3, 1 µM retinoic acid, 100 nM Tamoxifen)

4.2.2.3 Stimulation of pleural effusion cells

Tumour derived cells behave differently than healthy cells and as mentioned in chapter I "breast cancer" breast cancer cells also depend on the presence or absence of hormones. As primary breast cancer tissues were not available, pleural effusions of breast cancer patients were collected. The effusions contain different kinds of lymphocytes and tumour cells and are thus a valuable source for primary breast tumour cells. After depletion of macrophages the cells were cultured in 6 well plates and as soon as possible stimulated with different agents over a time period of 4-5 days. The "conditioned" cell culture medium contained cell free effusion supernatant and RPMI medium mixed in a 1:2 ratio.

The two controls (cells in culture without any additives (w.a.) versus cells after depletion of macrophages and snap frozen uncultured cells) of pleural effusion derived from breast cancer patients (HD-A-58 – HD-A-88) are compared regarding their NY-BR-1 expression (figure 4.13). The NY-BR-1 expression was normalized against the housekeeping gene ß-Actin.

Compared to controls of the tissues and isolated epithelial cells the opposite effect can be observed in the controls of pleural effusions: NY-BR-1 expression is lower in the uncultured cells than in the cultured cells in 6 out of 7 patients. In patient HD-A-66 no NY-BR-1 expression is detectable (C_P: 35) in both controls. The NY-BR-1 expression of non-cultured cells in patients HD-A-66 andHD-A-68 is not detectable (C_P: 35) but in cultured non-stimulated cells a weak expression (C_P: 31/30) can be measured. Patient HD-A-86 has no NY-BR-1 expression in w.a. cells (C_P: 35) but a very low expression in non-cultured (C_P: 34). The ß-Actin C_P-values range from 14 to 24 depending on the patient.

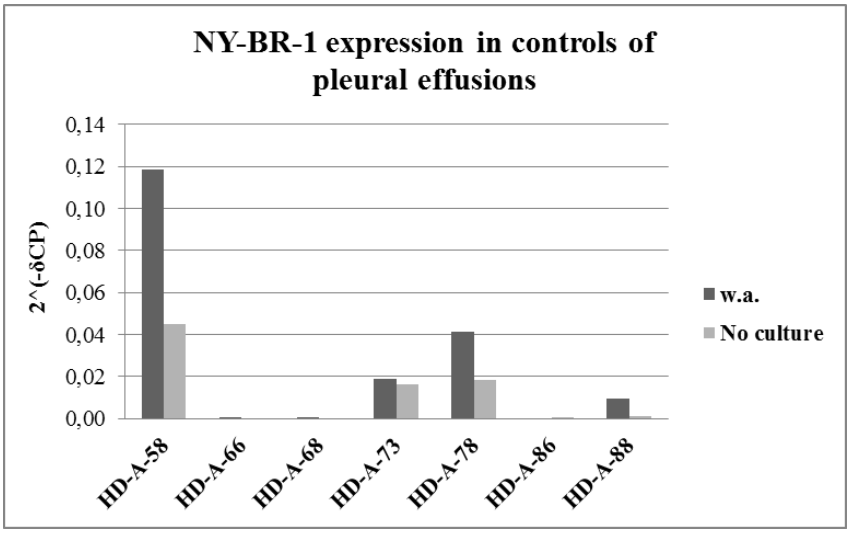


Figure 4.13: NY-BR-1 expression in the controls of pleural effusions derived from breast cancer patients at different stages of therapy. The dark grey bars (without additives (w.a.)) represent the NY-BR-1 expression in cells cultured as long as all cells but no supplements were given. The light grey bars show the NY-BR-1 expression in snap-frozen uncultured cells.

Figure 4.14A shows the results of the stimulation of seven different patients. The NY-BR-1 expression on RNA level in patients HD-A-66, -68 and -86 is in general very weak or not existing (C_P: 30-35). In patient HD-A-58 a high NY-BR-1 expression was observed (C_P: 20-24). 5'Aza (C_P: 21), Dex (C_P: 22), E2 (C_P: 21) and VitD3 (C_P: 21) up-regulate NY-BR-1 expression in patient HD-A-58, whereas progesterone (C_P: 22), retinoic acid (C_P: 22) and TSA (C_P: 20) down-regulate it compared to the control (w.a.). In patients HD-A-73 and HD-A-88 a weak effect induced by TSA (C_P: 23) can be observed, which down-regulates the NY-BR-1 expression. In patient HD-A-78 a clear down-regulation can be detected after adding Dex (C_P: 23), progesterone (C_P: 22), retinoic acid (C_P: 22) and TSA (C_P: 22) compared to the non-stimulated cells.

Also the pleural effusion cells were stimulated with 5'Aza in combination with all other agents (figure 4.14B). The combinations Aza/Dex (C_P: 21-24) and Aza/TSA (C_P: 22-24)

down-regulate the NY-BR-1 expression in patients HD-A-58, -78 and -88. All other combinations (Aza/E2 (C_P: 21), Aza/VitD3 (C_P: 20), Aza/Progesterone (C_P: 22), Aza/Retinoic acid (CP: 21)) induce an up-regulation of NY-BR-1 in patient HD-A-58 with Aza/Retinoic acid inducing the strongest effect. Aza/E2 (C_P: 21/22) and Aza/Progesterone (C_P: 23) seem to down-regulate the expression in two patients (HD-A-78 and HD-A-88). The cells of patient HD-A-73 do not show a definite effect after adding the combinations except for Aza/TSA (C_P: 23), which down-regulates the NY-BR-1 expression compared to the control. The combined stimulated cells of HD-A-66/68/86 just have a low NY-BR-1 expression (C_P: 30-35).

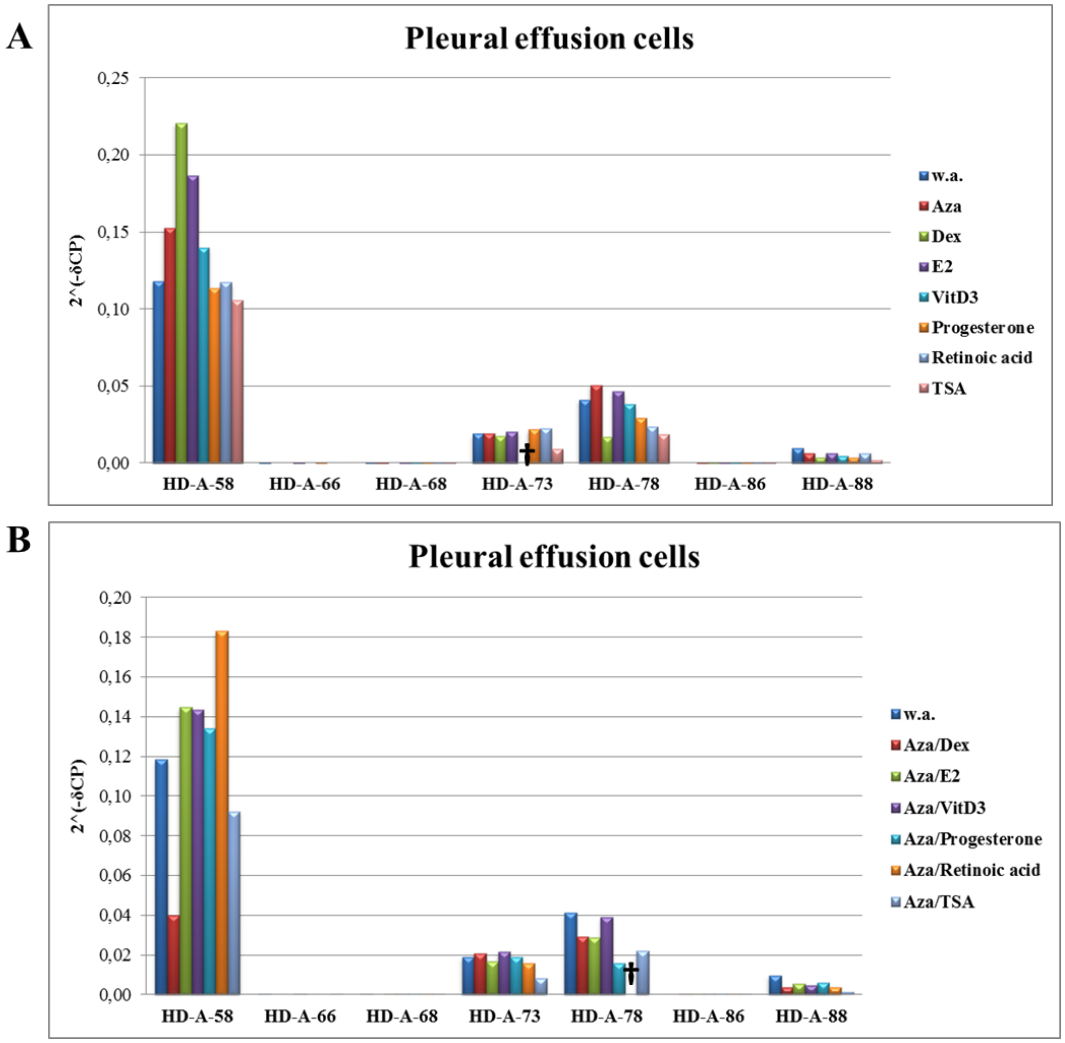


Figure 4.14: A: The effect on NY-BR-1 expression in pleural effusion cells derived from breast cancer patients at different stages of therapy. The non-stimulated cells functions as control (w.a.) B: The stimulation of pleural effusion cells with hormones combined with 5'Aza. (1 µM 5'Aza, 1 µM Dex, 10 nM E2, 1 nM VitD3, 200 nM TSA, 1 µM progesterone, 1 µM Dex, 1 nM VitD3, 1 µM retinoic acid) (†: no value available)

Due to a higher amount of obtained cells from four patients it was possible to test other stimulation agents like Tamoxifen and VPA, also in combination with 5'Aza (figure 4.15). All tested agents down-regulate NY-BR-1 expression compared to the control in patient HD-A-88 with Aza/VPA (C_P: 23) having the strongest effect. In all three patients (HD-A-73, -78, -88) a down-regulation could be detected after adding the HDAC inhibitor VPA (C_P: 22-23). Tamoxifen (C_P: 21), which binds to oestrogen receptor, and Aza/Tamoxifen (C_P: 22/21) results in an up-regulation of NY-BR-1 expression in two patients (HD-A-73 and HD-A-78). Aza/VPA does not affect NY-BR-1 expression in patient HD-A-78 (C_P: 21) and shows a weak effect in patient HD-A-73 (C_P: 22).

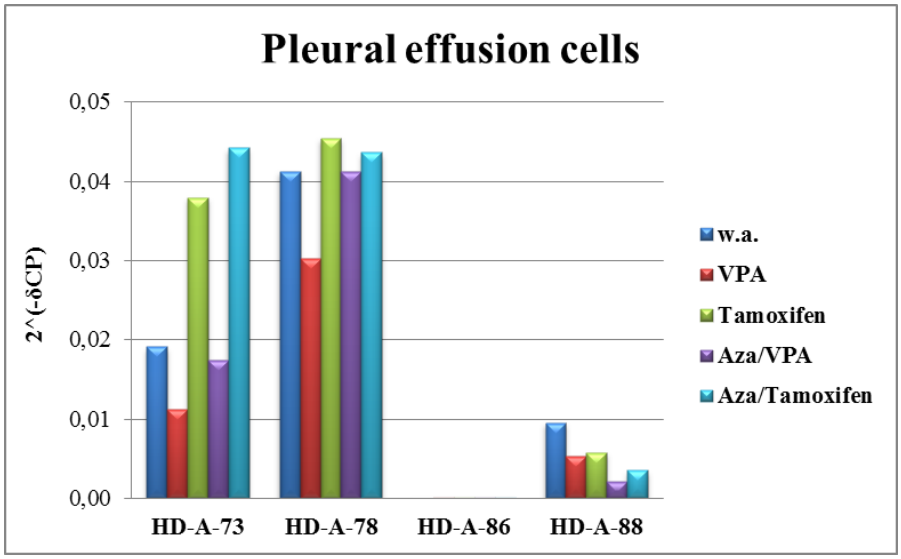


Figure 4.15: Results of the stimulation experiment of pleural effusion cells derived from breast cancer patients with additional substances. The non-stimulated cells function as control (w.a.). (1 µM 5'Aza, 1 mM VPA, 100 nM Tamoxifen)

Approximately two months later another pleural effusion from patient HD-A-88 was obtained (HD-A-90). Due to the high amount of cells time kinetics (6,8,10, 12 h) were performed with E2, progesterone, VitD3, retinoic acid, VPA, TSA and in combination (figure 4.16). The primary tumor was scheduled as triple positive (ER/PR/HER2). The highest up-regulation of NY-BR-1 compared to the controls can be achieved with E2 (6 h) (C_P: 21), progesterone (6, 8, and 12h) (C_P: 21, 23, 21), VitD3 (6, 8, and 12h) (C_P: 23, 23, 22), retinoic acid (6 and 12 h) (C_P: 23, 22), and VPA (6 and 12 h) (C_P: 24, 22). The combinations VPA/Progesterone (C_P: 24, 24, 24, 22), VPA/VitD3 (C_P: 24, 23, 23, 22) seem to up-regulate NY-BR-1 expression throughout the whole time period but to a lower extent than the single applied hormones. Many combinations with TSA (TSA/E2, VPA/TSA/E2, VPA/TSA/Progesterone, TSA/VitD3,

VPA/TSA/VitD3, VPA/TSA/Retinoic acid) seem to inhibit NY-BR-1 expression at specific time points.

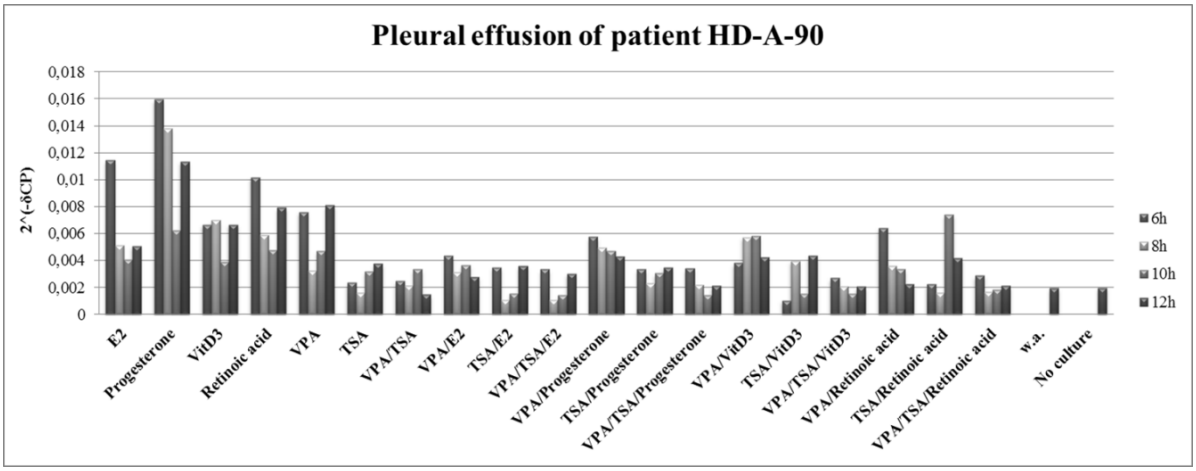


Figure 4.16: The effect on NY-BR-1 expression in pleural effusion cells derived from a triple positive breast cancer patient after stimulation with different agents. A time kinetic was performed to analyze whether the effect on NY-BR-1 expression will change over time after adding specific hormones and HDACs inhibitors. The non-stimulated (w.a.) and uncultured (no culture) cells function as controls.

4.2.2.4 Characterization of pleural effusion cells

As described in chapter 1.4 breast tumours are characterized by different molecular subtypes and due to these subtypes therapy needs to be adjusted. The pleural effusion cells were obtained from breast carcinoma patients and thus they have a specific molecular subtype pattern. In table 4.3 the characteristics of the primary tumours is summarized.

Table 4.3: Summary of the characteristics of the primary tumours of the patients whereof the pleural effusions were obtained. The breast cancer classification is done according to TNM staging. The receptor status is calculated either with IRS score, Allred score or given in percent. (ID: Identification, T: Tumor, N: Lymph node, M: Metastases, G: Grade, -: negative, NA: not available)

ID	Diagnosis	Material	T	N	M	G	ER	PR	Her2	p53	bcl-2	Ki 67	Histological type
HD-A-58	MaCa 2006	Ascites	4	3	0	2	-	-	-	NA	NA	NA	invasive lobular
HD-A-66	MaCa 2004	Pleural effusion	4	3		2	90%	30%	0	0	0	5%	invasive ductal
HD-A-68	MaCa 2010	Ascites	2	0	0	2	12	12	-	NA	NA	NA	invasive ductal
HD-A-73	MaCa	Ascites	3	3	1	2	12	6	2+	NA	NA	NA	invasive lobular
HD-A-78	MaCa	Ascites	2	3	1	NA	9	6	-	NA	NA	10%	invasive lobular
HD-A-86	MaCa	Ascites	NA	NA	NA	NA	0/8	3/8	-	0	0	60%	invasive ductal
HD-A-88	MaCa	Ascites	2	0	1	NA	100%	100%	1+	5%	3	30%	invasive ductal

In FACS analysis α-NY-BR-1#2 (mouse) antibody was used for intracellular (IC) and extracellular/surface (EC) staining, a surface staining was done for EpCAM and for HER2 and an IC staining with mouse α-human Cytokeratin 14, 15, 16, and 19-PE. Here, exemplary results are shown for three different patients (figure 4.17). In all measurements the red line represents the isotype control and the blue line indicates the results of particular receptor/protein of interest. A gate was set in the FSC/SSC setting to get the tumour cells for further analysis (figure 4.17A). In figure 4.18B the results for the CK (EC) staining are shown and HD-A-73/88 are negative for all tested CK and a small population is positive for CK in patient HD-A-68. All three patients are positive for HER2 (figure 4.17C) and except patient HD-A-68 also for EpCAM (figure 4.17C/D). In figure 4.18E/F the results of the NY-BR-1 staining is shown and all three patients are NY-BR-1 negative on the surface but patient HD-A-73 and HD-A-88 are intracellular positive for NY-BR-1. HD-A-68 is completely NY-BR-1 negative.

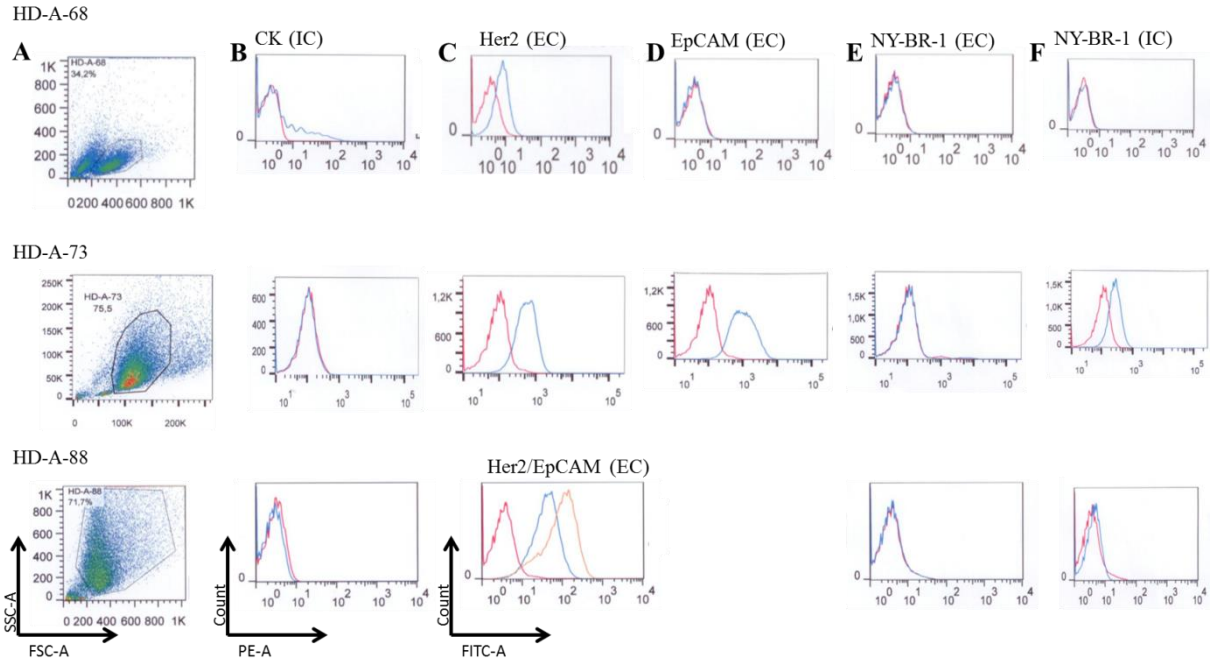


Figure 4.17: Results of the FACS analysis of pleural effusion cells of three patients. A: A gate was set to receive the tumour cell population for further analysis. B: Intracellular cytokeratin 14, 15, 16, and 19 staining. C: Patient HD-A-68/73: Cell surface Her2 staining, Patient HD-A-88: Her2 (orange) and EpCAM (blue) staining. D: Patient HD-A-68/73: cell surface EpCAM staining. E: Cell surface NY-BR-1 staining. F: Intracellular NY-BR-1 staining. (red line: appropriate isotype, blue line: results of the staining of the protein/receptor of interest.

Due to therapy applications and time the expression of the receptor can change. Therefore, the pleural effusion cells were characterized again by qPCR analysis. The Her2, ERα, PR and

integrin-α 6 expressions were normalized against the housekeeping gene ß-Actin. The C_P values of ß-Actin range in these analyses between 14 and 21. Integrin-α 6 was chosen because this integrin is a progenitor marker of basal/bipotent cells in human breast cells (Stingl et al., 2006). The C_P values of HER2 range between 16 and 27 (figure 4.18A), for ERα between 30 and 35 (figure 4.17B), for PR between 28 and 35 (figure 4.18C), for integrin-α6 22 and 31 (figure 4.18D). The HER2 expression is in some samples (HD-A-88/90) the highest compared to the other samples but also compared to ERα, PR and intgrein-α 6 expression. The missing samples of ERα (HD-A-86, -88, -90), PR expression (HD-A-86, -88, -90) and integrin-α 6 (HD-A-78) could not be analysed due to poor melting curves.

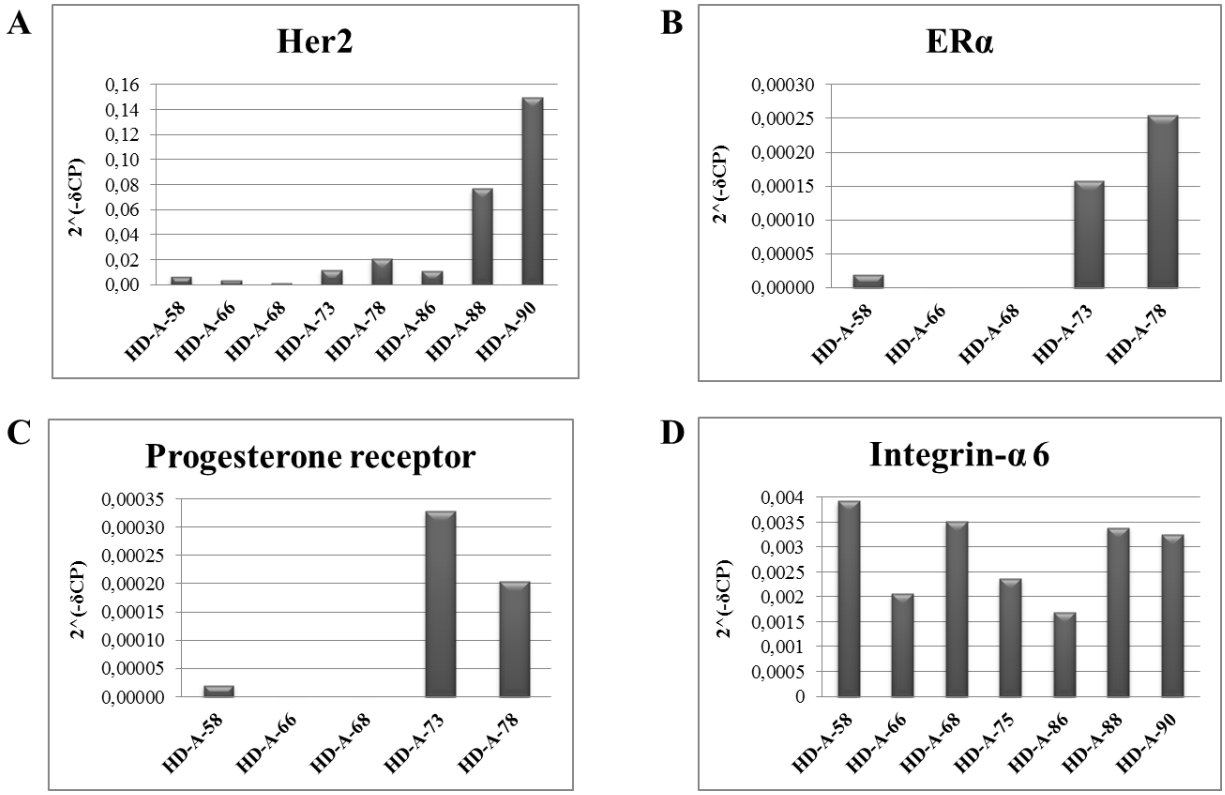


Figure 4.18: Receptor status analysis in pleural effusion cells by qPCR. A: Pleural effusion cells of eight patients were tested regarding their Her2 expression level. B: Pleural effusion cells of five patients were tested regarding their ERα expression level. C: Pleural effusion cells of five patients were tested regarding their PR expression level. D: Pleural effusion cells of seven patients were tested regarding their integrin-α 6 expression level. (Patient HD-A-75 correlates to patient HD-A-73)

In transiently transfected cell lines NY-BR-1 expression is lost after three to four days (figure 4.24). Therefore, it was interesting to analyse how quickly NY-BR-1 expression is down-regulated or whether it can be maintained by culturing with conditioned medium in pleural effusion cells.

The cells of patient HD-A-26 were cultured for at least four weeks with conditioned medium and the NY-BR-1 expression on day zero, 11, 14, 18, 22 and 28 was analysed by qPCR. The C_P value at day zero is 20 and decreases to 23 at day 28 (figure 4.19). The ß-Actin C_P values range between 14 and 15.

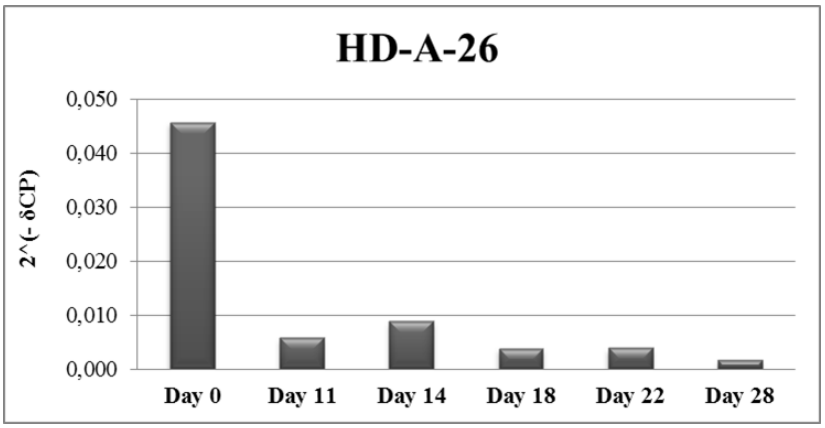


Figure 4.19: Time kinetic of pleural effusion cells of patient HD-A-26 analysed by qPCR. The cells were tested regarding their NY-BR-1 expression.

Summarizing these results, it is important to consider that *in vitro* culturing affects the NY-BR-1 expression in tissue pieces, isolated epithelial cells and pleural effusion cells. The effect on NY-BR-1 expression of a specific agent is depending on the setting: healthy vs. tumour cells or hormone sensitive vs. hormone negative. In general it could not be demonstrated that a substance generally leads to an up- or down-regulation of NY-BR-1 expression.

In table 4.3 the characterization of the primary tumour of each patient is summarized to give an overview of the grading and receptor status. The qPCR data of the pleural effusion cells showed that the receptor status, analysed on RNA level, has changed compared to the analysis of the primary tumour. NY-BR-1 expression can be mainly detected intracellulary via FACS analysis and time kinetics showed that NY-BR-1 is detectable over several weeks when cultured with conditioned medium.

4.2.2.5 Methylation status analysis

Methylation in the promoter region is an important regulatory mechanism in the transcription machinery. High methylation in a promoter region blocks transcription activation, whereas methylation downstream of the transcription start site enhances gene activation (Armstrong, 2014).

The methylation pattern can alter the transcription initiation and thus it was investigated whether a specific methylation pattern can be detected and correlated to NY-BR-1 expression

in stimulated samples. DNA was isolated from selected stimulated samples and bisulfit conversion was done. The methylation status of various fragments in the promoter region of NY-BR-1 was analysed using the EpiTYPER – MassARRAY technology in collaboration with the group of C. Gerhäuser, German Cancer Research Center. Amplicons were designed up to 1300 bp upstream from the transcription start site. As the promoter region has many repetitive elements, only six amplicons (number 53-58) could be designed and used for analysis (figure 4.20). Amplicons 55-57 are located in the CpG islands of the NY-BR-1 promoter.

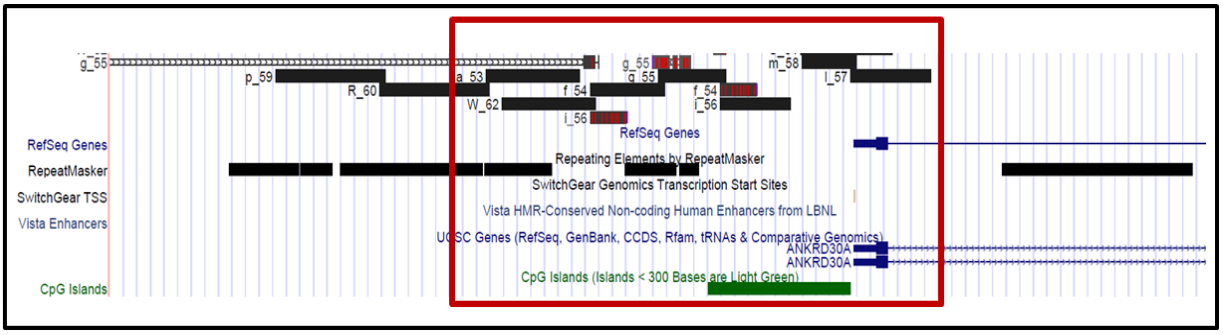


Figure 4.20: Overview of the designed amplicons with their binding sites. Amplicons number 54-58 with the full black bar were used for analysis.

The methylation status of several stimulated samples of pleural effusion cells (HD-A-73 and HD-A-78) and isolated epithelial cells (HD-T-292) were analysed in various amplicons in the promoter region of NY-BR-1 (figure 4.21). The median amplicon methylation was calculated and single dots represent different experimental settings. They are connected by lines for a better classification. Values on the y-axis indicate the percentage of methylation with one representing a 100 % methylation.

Interestingly, the amplicon 58, which ranges from the promoter over to the transcription start site, is highly methylated (median 1.0) in all samples except for patient HD-A-73 after Aza/TSA (median 0.4) stimulation. NY-BR-1 expression is decreased in this sample compared to the control cells as the qPCR results show (figure 4.14). In amplicon 57 differences between each patient become visible. All patients show a lower methylation compared to amplicon 58 (median around 0.9). Patient HD-A-73 has a lower methylation status after TSA application (median 0.8) but an increase after VPA (median 1.0) stimulation in amplicon 57. In both cases the NY-BR-1 expression is lower compared to the control cells. In patient HD-A-78 an increase of methylation is observed after 5'Aza stimulation (median ~ 0.98) compared to the control cells, which is striking because 5'Aza is a demethylation agent. The combination Aza/VPA decreases the methylation (median 0.8) in patient HD-A-78 in

amplicon 57. The NY-BR-1 expression is higher with 5'Aza stimulation and with Aza/VPA it stays the same. The isolated epithelial cells of patient HD-T-292 have a methylation increase after Aza/VPA) stimulation (median ~0.98). All samples detected with amplicon 56 have a decreased methylation pattern compared to all other amplicon areas (median between 0.2 and 0.4). In patient HD-T-292 the methylation is decreased by Tamoxifen stimulation (median 0.1) but slightly increased with Aza/VPA (median 0.25). Also in patient HD-A-78 a weak increase of methylation can be observed with Aza/VPA (median 0.4) stimulation in amplicon 56 compared to the control cells (median 0.29). After TSA or Tamoxifen stimulation a decreased methylation pattern of approximately 50 % (meodian 0.5) is observed in patient HD-T-292 in the area of amplicon 55. Patient HD-A-78 has an increased methylation after Aza/TSA (median 0.9) stimulation in amplicon 55. In amplicon area of 54 a demethylation occurs by TSA stimulation in patient HD-A-73 but not in the other two patients. Tamoxifen induces a demethylation (median ~ 0.6) in patient HD-A-78 compared to the control cells (median 0.9).

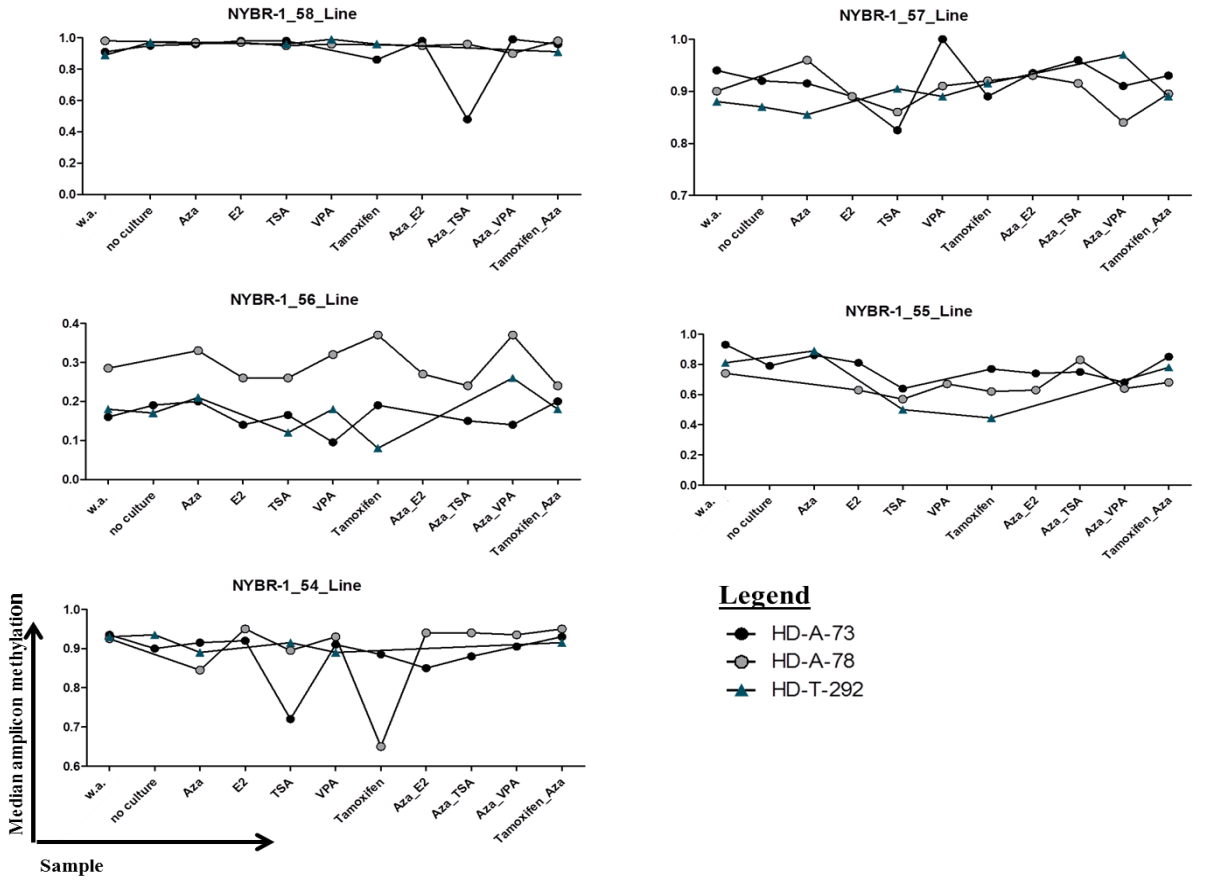


Figure 4.21: Methylation status analysis of stimulated pleural effusion cells (HD-A-73/78) and isolated epithelial cells (HD-T-292) of various fragments in the NY-BR-1 promoter region up to 1300 bp upstream of transcription start site. Median amplicon methylation values were calculated for all samples (1=100%).

To summarize this part of the results: no unique methylation pattern can be related to NY-BR-1 expression after a certain stimulation agent was applied. The effects on NY-BR-1 up- or down-regulation by an agent vary from patient to patient and from cell subset to cell subst. Moreover, in pleural effusion cells the receptor status of the cells self and not compulsory the receptor status of the primary tumour needs to be considered.

4.3 The effect of NY-BR-1 over-expression on proliferation and on the cell cycle

4.3.1 Proliferation assay

Preliminary data from soft agar assays (Group of Professor Jäger) revealed that transient over-expression of NY-BR-1 in contrast to controls leads to colony formation suggesting that NY-BR-1 may play a role in tumourgenesis *in vitro.*

To test the effect of NY-BR-1 over-expression in more detail, proliferation assays were performed and cells were analysed by flow cytometry. In contrast to soft agar assays, the transfection efficiency can be determined very quickly, the analysis of all settings is fast, single cell populations can be analysed, NY-BR-1^+ cells can be discriminate from NY-BR-1- cells in one cell population, different kind of settings can be applied at once (e.g. proliferation, apoptosis).

Four different cell lines, the normal epithelial embryonic kidney cell line HEK293, and the normal epithelial breast cell line MCF-10A as well as the epithelial embryonic kidney, containing the SV40 large T-antigen, cell line HEK293T and the breast cancer cell line MCF-7 were transiently transfected with the pc3 NY-BR-1 GFP plasmid. The cells were labeled with the proliferation marker PKH26, which stably incorporates a yellow orange fluorescent dye with long aliphatic tails into lipid regions of the cell membrane (Wallace et al., 2008). During cell proliferation the PKH26 dye gets distributed equally between daughter cells and the intensity decreases with each division. The cells were kept in culture over a time period of one to six days and prior to flow cytometric analysis, they were labeled with 7-AAD and Annexin V APC to determine the living/dead and apoptosis ratio. Non-transfected and pc3 GFP transfected cells served as controls. The evaluation was done with the software Flowjo 7.6.5. Exemplary results are shown for the two cell lines HEK293 and MCF-7, results for MCF-10A and HEK293T cells are in the appendix (figure 6.3-6.6).

The following settings for the flow cytometric analysis were applied: forward (FSC)/sideward (SSC) scatter to discriminate the cells due their size and density, PKH26 (PE)/GFP (FITC) to separate the cells into GFP positive/negative. The next step was to analyse the proliferation

profile of each population on all days with PKH26 (PE)/Count in a histogram. The last setting was Annexin (APC)/7-AAD to detect the amount of live/dead and apoptotic cells (see 3.2.3.5).

The first gate comprised of almost the whole cell population, just leaving the cell debris out (figure 4.22A). This population was divided into GFP^+ and GFP^- cells (figure 4.22B) and analyzed for PKH26 fluorescence (figure 4.22C). The cells were labeled with 7-AAD, a live/dead cell marker, and Annexin V, an apoptosis marker (figure 4.22D). The left lower quadrant contains the cells, which are 7-AAD and Annexin V negative. The left upper quadrant represents the cells, which are 7-AAD positive but Annexin V negative. The right upper quadrant shows the double positive cells-cells, which are dead through apoptosis. The 7-AAD negative but Annexin V positive cells are shown in the right lower quadrant.

The PKH26 labeling decreases from day one to day six and this indicates that the untransfected HEK293 cells (GFP^-) proliferate. The main cell population is 7-AAD/Annexin V negative (80-89 %) in all checkpoints. The double positive cell population decreases from 6 % (day one) to 0.7 % (day six), therefore the 7-AAD cell population increases from 3.9 % to 9.6 %. The apoptotic cell rate (Annexin V positive) decreases from 3.9 % to 0.8 %. The untransfected cells are vital and just a small part of the population is dying due to apoptosis.

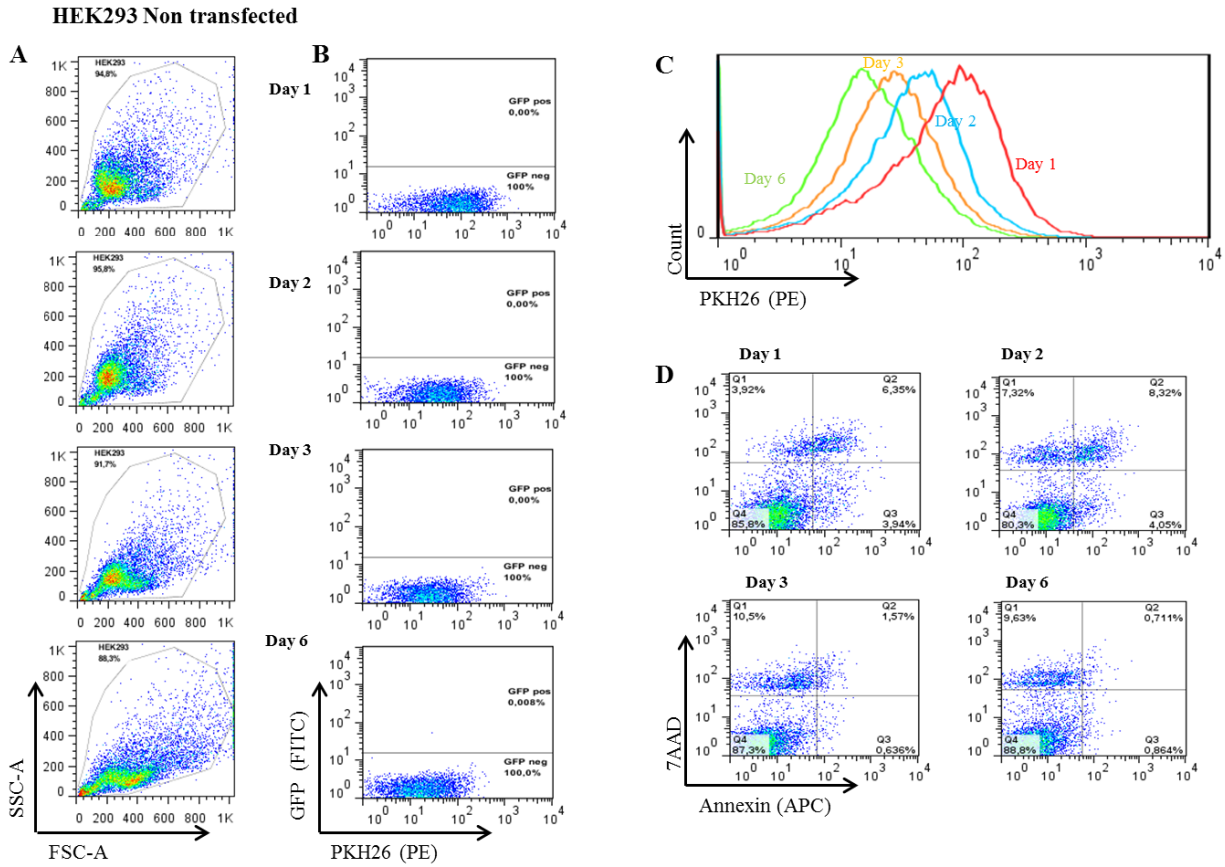


Figure 4.22: Overview of the proliferation assay of untransfected HEK293 cells. A: Selected cell population for further analysis leaving without cell debris. B: The cell population divided into GFP positive/negative cells for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

The same was done with the pc3 GFP transfected HEK293 cells as described above. The GFP transfected HEK293 cells are here shown exemplary for all other cell lines. The GFP^- cell population was analyzed independently from the GFP^+ population. The GFP^+ population decreases over the measured time from 22.6 % to 15.7 % (figure 4.23B). The PKH26 amount decreases as well in both populations leading to the conclusion that both cell populations proliferate (figure 4.23C). The main GFP^- population is double negative ranging from 80.9 % to 87.7 %. The amount of double positive cells decreases from 10 % to 0.6% on day six. The same can be observed in the GFP^+ population with 69 % to 95 % double negative for 7-AAD and Annexin V and a decrease of the double positive cells from 4 % on day one to 0.1% on day 6 (figure 4.23D). The results show that the transfection itself had no influence on the proliferative manner of the cells and is not increasing the apoptosis rate.

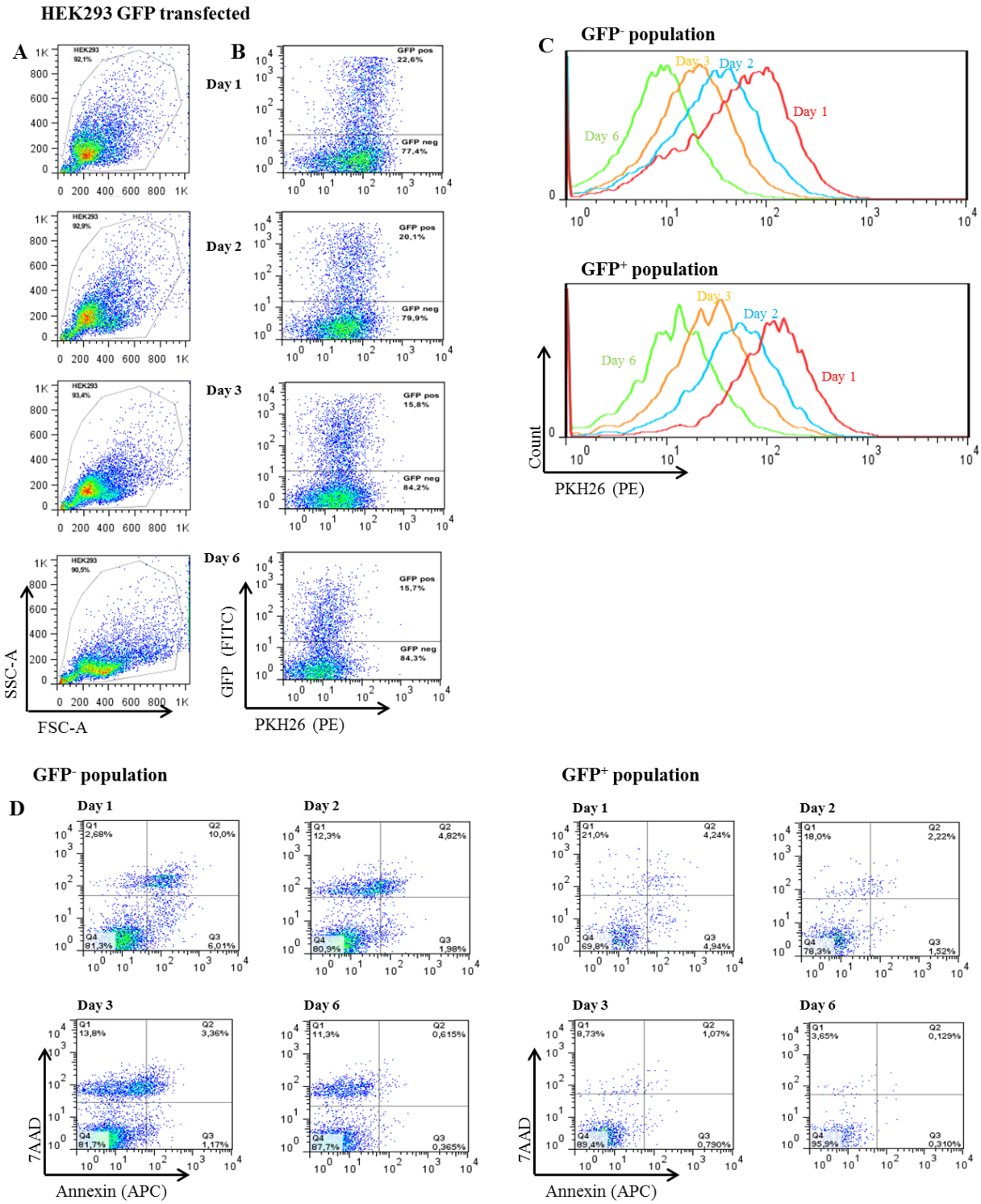
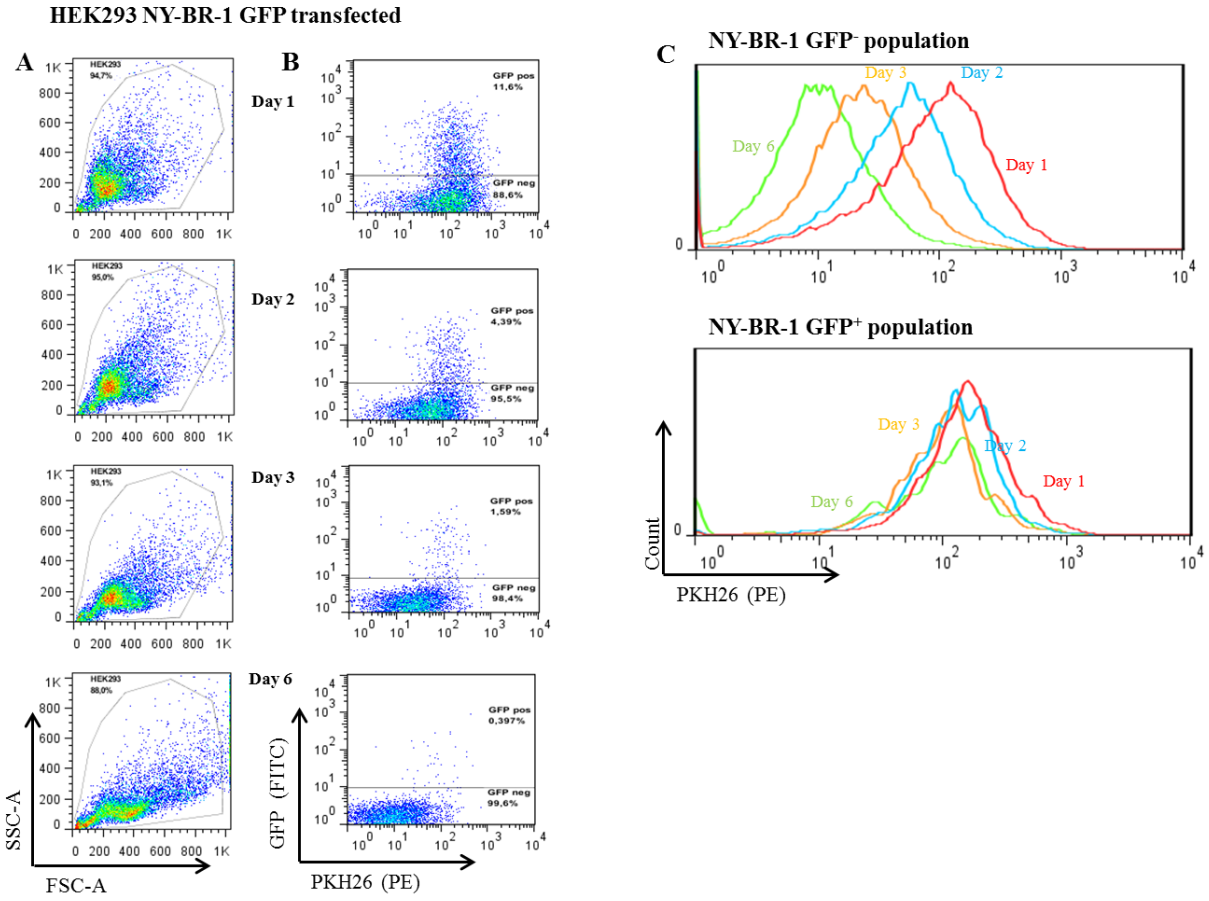


Figure 4.23: Overview of the proliferation assay of HEK293 pc3GFP transfected cells. A: Selected cell population for further analysis leaving without cell debris. B: The cell population divided into GFP positive/negative cells for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

The HEK293 cells were also transfected with pc3 NY-BR-1 GFP and analysed the same way as the untransfected and pc3 GFP transfected cells. Although the NY-BR-1 positive cells decrease rapidly, starting on day one with 11.6 % ending with 0.39 % on day six (figure 4.24B), the proliferation profile of both populations (NY-BR-1 negative vs. positive) could be analysed. The NY-BR-1 GFP negative population shows the same proliferative profile as the untransfected and GFP transfected cells: they proliferate over throughout the whole observation time. The NY-BR-1 GFP positive population does not proliferate at all (figure 4.24C). Both cell populations are mainly 7-AAD/Annexin negative. 7.7 % of the NY-BR-1 GFP negative population is double positive with decreasing incidence until day six (0.48 %). The NY-BR-1 GFP positive population contains 1.6 % double positive cells on day one and declines to 0.9 % on day six. It is worth to mention that the 7-AAD positive cell population increases over time in both populations (NY-BR-1 GFP^-: 3.9 -15.5 %, NY-BR-1 GFP^+: 0.4 – 4.7 %), whereas the Annexin V positive population decreases (NY-BR-1 GFP^-: 6.9 -0.3 %, NY-BR-1 GFP^+: 8.8 – 1.9 %) (figure 4.24D).



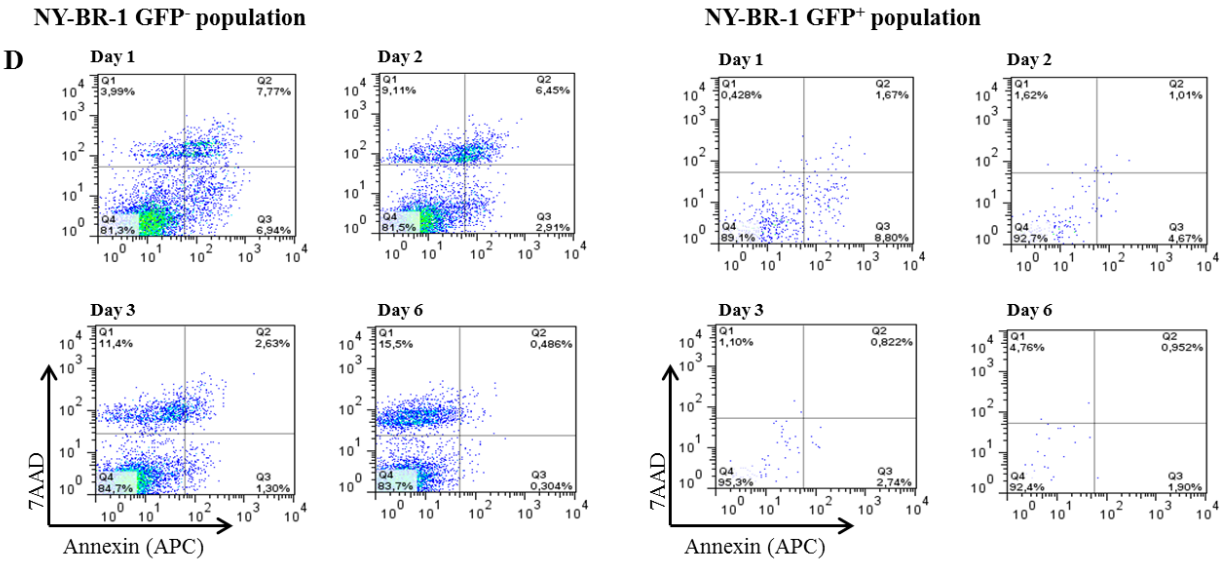


Figure 4.24: Overview of the proliferation assay of HEK293 pc3NY-BR-1 GFP cells. A: Selected cell population for further analysis leaving without cell debris. B: The cell population divided into GFP positive/negative cells for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

As NY-BR-1 is over-expressed in approximately 70 % of all breast tumors, the effect of NY-BR-1 over-expression on the proliferation was also analysed in the breast cancer cell line MCF-7.

In general, the MCF-7 cell line is growing slower compared to the HEK293 cell line in our culture system. The measured cells were gated to leave the cell debris out (figure 4.25A) and then divided into $GFP^{+/-}$ population for further analysis (figure 4.25B). In figure 4.25C the untransfected MCF-7 cells loose the proliferation marker PKH26 from day one to day six indicating that they proliferate. The main population can be found in the double negative compartment at all times but a lot of cells are also double positive ranging between 33.2 % on day one, dropping to 3.8% on day two but increasing to 28.6 % on day six. The single 7-AAD positive population rises up from 5 % (day one) to 22 % (day six) but the Annexin V positive population decreases from 7.9 % (day one) to 2.4 % (day six) (figure 4.25D). Taken these results together, the MCF-7 untransfected cells proliferate but due to culturing or labeling a higher population is apoptotic or dead compared to the HEK293 non-transfected cells.

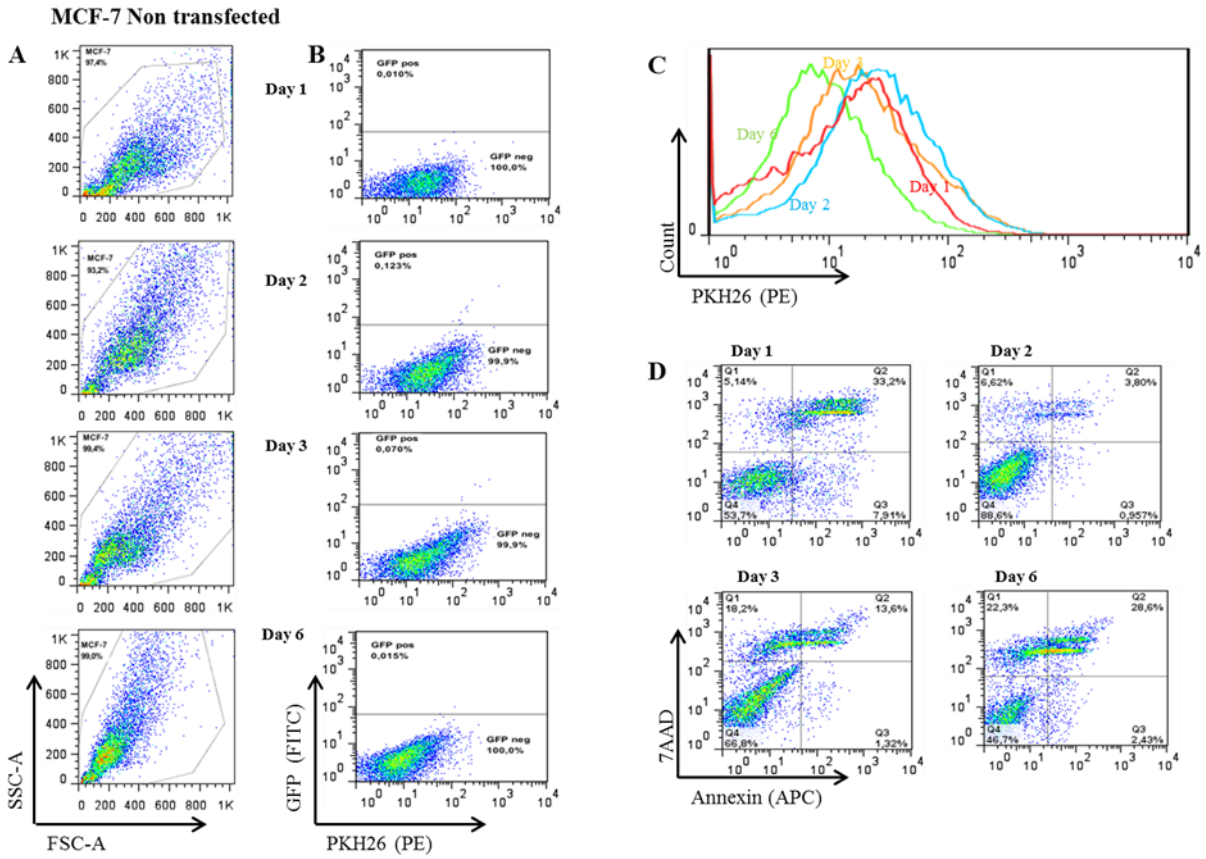


Figure 4.25: Overview of the proliferation assay of untransfected MCF-7 cells. A: Selected cell population for further analysis leaving without cell debris. B: The cell population divided into GFP positive/negative cells for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

Breast cell lines are in general difficult to transfect. That is the reason why the transfection efficacy is low (16.7 % day one) and NY-BR-1 GFP expression is lost during the measurements (1.89 % day six) (figure 4.26B). The NY-BR-1 GFP negative cell population proliferates but the NY-BR-1 GFP positive population does not (figure 4.26C). The 7-AAD/Annexin V staining of the NY-BR-1 GFP negative population shows that a lot of cells are double negative (37.9 – 52.9 %). The amount of the double positive cells drops down starting at 32.4 % on day one and ending at 23.3 % on day six. The 7-AAD cell population ranks around 20 % on all days, whereas the Annexin V positive population of the NY-BR-1 GFP negative cells decreases from 11.2 % (day one) to 3 % (day six). The 7-AAD/Annexin staining of the NY-BR-1 positive cell population shows that the double positive cells increase from 15 % (day one) to 23.8 % (day six) while dropping down in between to 8.4% on day three. Also the single positive cells for 7-AAD or Annexin V increase from day one to day six but decrease on day two and three (figure 4.26D).

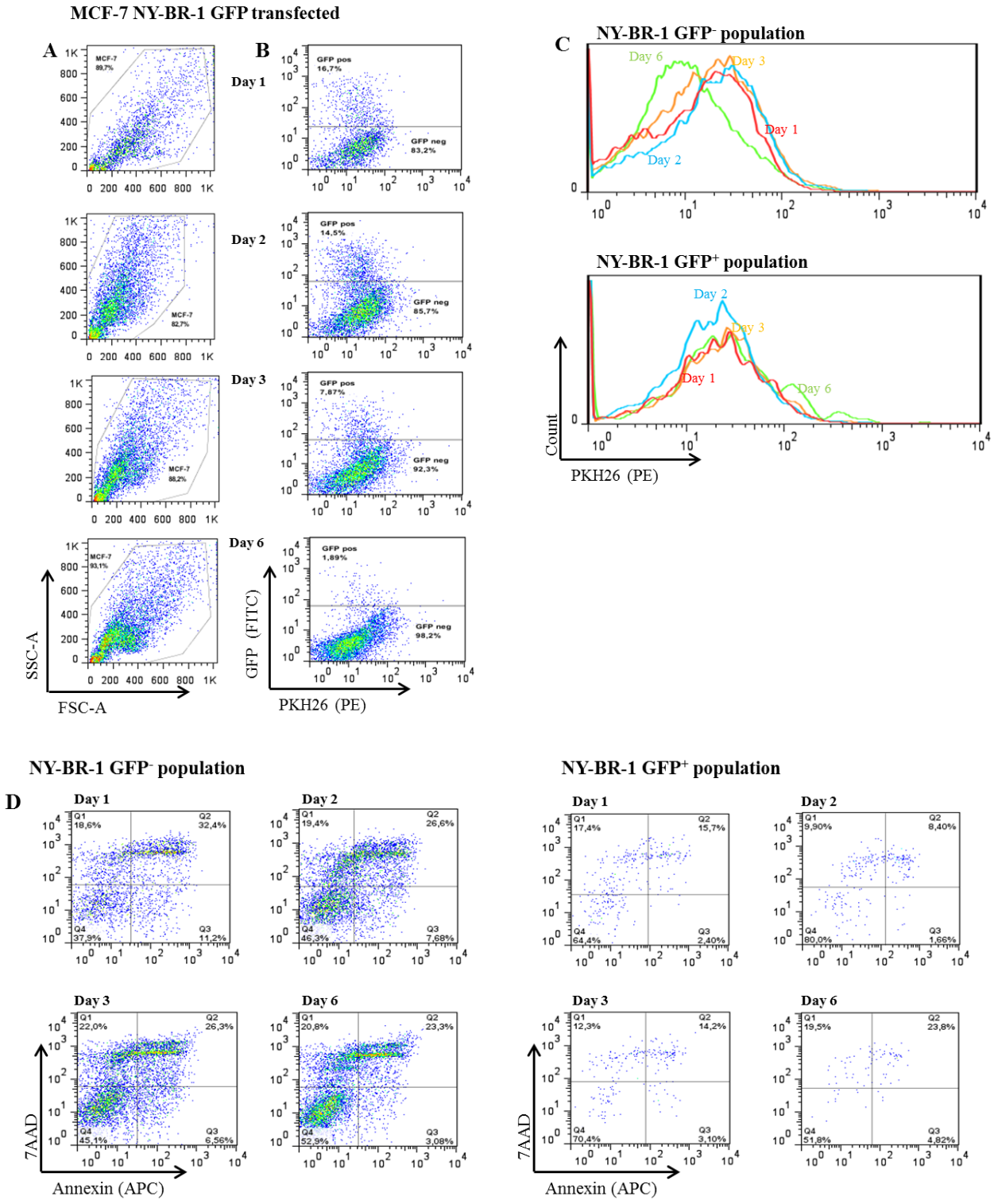


Figure 4.26: Overview of the proliferation assay of MCF-7 pc3NY-BR-1 GFP cells. A: Selected cell population for further analysis leaving without cell debris. B: The cell population divided into GFP positive/negative cells for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

The results of all four cell lines show that the non-transfected, GFP transfected and NY-BR-1 GFP transfected but being NY-BR-1 GFP negative cells are proliferating compared to the NY-BR-1 GFP positive cells. Moreover, the double positive cells (7-AAD/Annexin V) decrease over time except for in MCF-7 cells.

4.3.2 Cell cycle analysis

The cell cycle can be divided into the interphase (cell prepares for cell division), the mitotic phase (cell splits in two cells) and cytokinesis (describes the fully divided cells). In cell cycle analysis by flow cytometry the differences of DNA content are measured. The DNA content reveals information about cell ploidy, cell position in the cell cycle and shows the amount of apoptotic cells because their DNA is fractional. As shown in figure 4.27 four different phases can be discriminate in cell cycle analysis. The sub G1 peak ("sub-diploid") represents the dead cells due to the fact that apoptotic cells have fragmented mono- and oligo-nucleosomal DNA. In normal dividing cells the G0/G1 peak shows the cells in the pre-replicative cells. At this point the biosynthesis rate will be the highest and the duration of how long a cell stays in this phase can vary from cell to cell. During the S phase the DNA replication actually starts and the cells checks whether it is ready to enter the mitotic phase (G2 +M) (Darzynkiewicz et al., 2010).

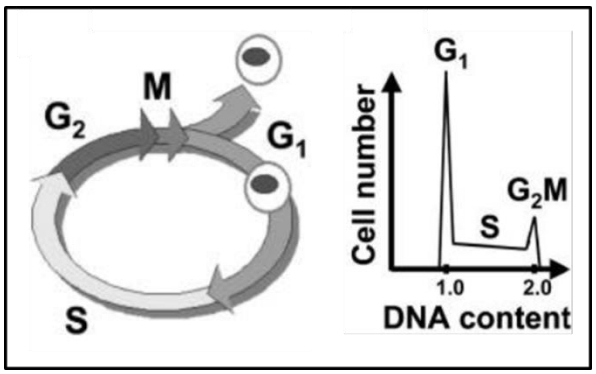


Figure 4.27: Schematic representation of the unique cell cycle phases (left) and its display in cell cycle analysis by flow cytometry (right). (adapted from (Darzynkiewicz et al., 2010) *)*

The dye PI was used to analyse the cells. PI intercalates with the DNA bases but also with RNA. Therefore, RNase was added after EtOH fixation to ensure that only the DNA content is analysed (see 3.2.3.3).

The results of the proliferation assay showed no proliferation of NY-BR-1^{+} cells and also almost no increase in dead or apoptotic cells. Therefore, a cell cycle analysis of all four cell lines (HEK293, HEK293T, MCF-7, and MCF-10A) was performed to analyse whether these

cells undergo G1 phase arrest. Here, two cell lines (HEK293, MCF-7) are shown exemplary. The other two analysed cell lines can be found in the appendix (figure 6.7-6.10)

The cells were transiently transfected with either pc3-NY-BR-1-GFP or pc3 GFP plasmid and fixed after 24 h with ice cold EtOH. RNase digestion was done followed by a staining with PI (living/death marker). The cells were analysed by FACS and Flowjo. Non-transfected cells function as a control. The pc3 GFP transfected cells could not be analysed because the GFP was washed out during EtOH fixation due to the small size of the protein.

The whole cell population (figure 4.28A) was separated into GFP positive and GFP negative cells (figure 4.28B). A gate was set (PI-Area vs. PI-Width) on the discriminated cell population to leave duplets and cell debris out (figure 4.28C). The accumulation of the cells in the single phases is shown in figure 4.28D of all three days. Between 2.84 % and 4.49 % of the HEK293 non-transfected cells can be found in the sub G1 phase. Around 50 % of the analysed cells are detected in the G1 phase on all three days. Between 9.9 % and 14.9 % of the cells prepare themselves for DNA replication and are represented in the S phase. The amount of cells accumulating in the G2 phase varies between 26.6 % and 37.3 %.

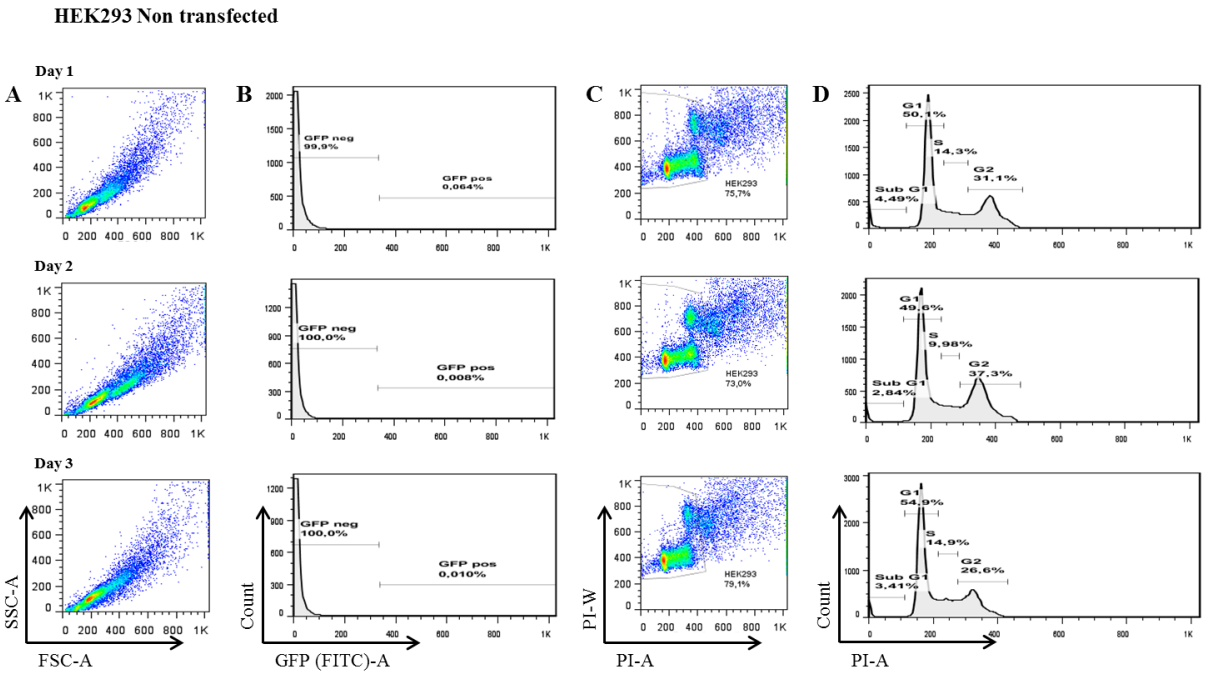


Figure 4.28: Results of the cell cycle analysis of HEK293 non-transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell debris out. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

The same analysis was carried out with the pc3 NY-BR-1 GFP transfected cells. Here some differences are revealed compared to the non-transfected cells. First of all the measured cells were divided into NY-BR- GFP positive vs. negative cells (figure 4.29B). The gate was set to leave the duplets and cell debris out (figure 4.29C) and with this population the cell cycle analysis was done.

The amount of cells in the sub G1 phase of the NY-BR-1 GFP negative population decreases from 13.2 % (day one) to 5.28 % (day three). Around 55 % of this population accumulate in G1 phase during all three measuring points. Between 4.36 % and 10.4 % of the cells are in S phase and between 23.1 % and 32.5 % of the cells are found in G2 phase. The cells present in the sub G1 phase of the NY-BR-1 GFP positive cell population increase from 1.14 % (day one) to 3.36 % (day three). Compared to the NY-BR-1 GFP negative cells more cells of the NY-BR-1 GFP positive population accumulate in G1 phase (between 67.1 % and 74.4 %). Approximately 4 % of the NY-BR-1 GFP^+ cells are in S phase and about a fifth can be found in the G2 phase (figure 4.29D).

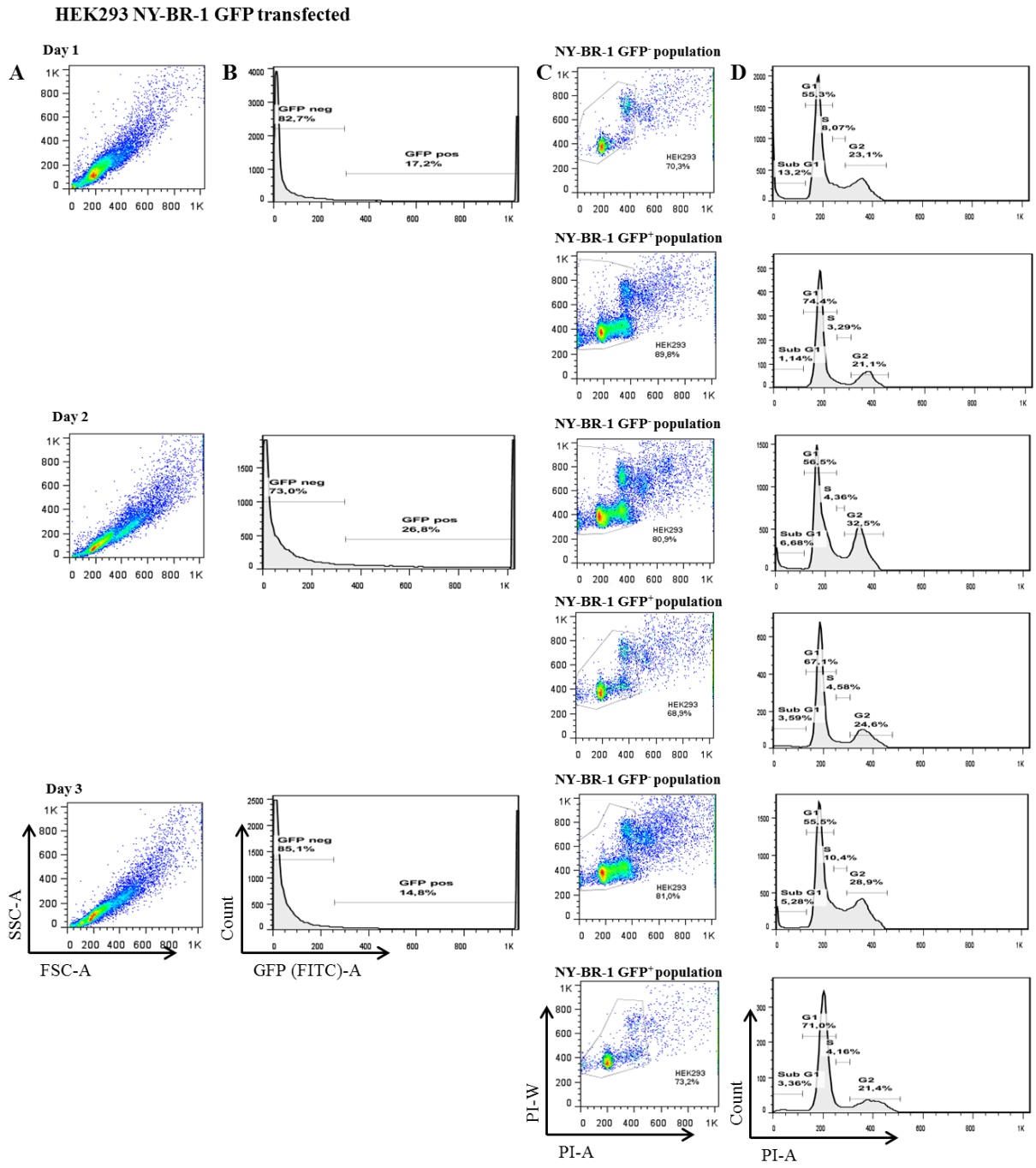


Figure 4.29: Results of the cell cycle analysis of HEK293 NY-BR-1 GFP transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell debris out in either NY-BR-1 GFP negative or NY-BR-1 GFP positive cell population. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

The second analysed cell line was the breast cancer cell line MCF-7. The cells were treated the same way as HEK203 cells with measuring points on day one, two and three. The whole cell population (figure 4.30A) was divided into GFP positive or negative populations (figure 4.30B). The non-transfected cell population is GFP negative and in figure 4.30C a gate was set to get the duplets out for further cell cycle analysis (figure 4.30D).

The amount of cells found in subG1 phase rises from 13.1 % (day one) to 27.2 %. The numbers of cells in the G1 phase decrease, starting at 63.4 % on day one and being at 47.5 % at day three. Cells, which replicate their DNA (S phase), increase from 2.54% (day one) to 4.18 % (day two) to 4.92 % (day three). Approximately 20 % of the analysed cells accumulate in G2 phase, with a drop down on day two to 15.7 %.

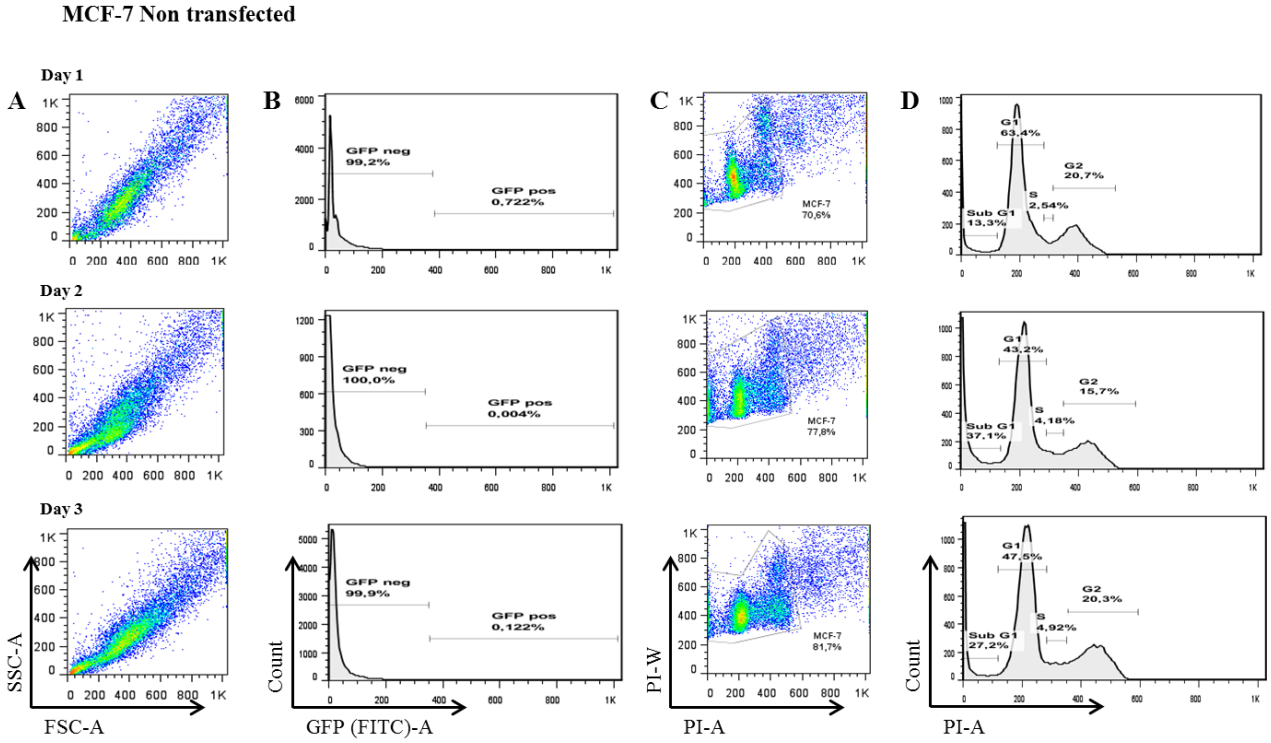


Figure 4.30: Results of the cell cycle analysis of MCF-7 non-transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell debris out. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

The MCF-7 NY-BR-1 GFP transfected cells were divided into NY-BR-1 GFP positive or negative populations (figure 4.31B). A gate was set on each cell populations to leave cell debris and duplets out for further analysis (figure 4.31C). 16 % of the NY-BR-1 GFP^- population can be found back in the sub G1 phase on day one with increasing amount on day three until 30 %. The amount of cells in the G1 phase decreases from 61.2 % on day one to 50.5 % on day three. Therefore the cells in the S phase increase from 0.81 % to 2.28 % (day one to three). Less cells of the NY-BR-1 GFP^- population are present in the G2 phase: starting at 22 % (day one) and dropping down to 17.2 % (day three). More cells of the NY-BR-1 GFP^+ population accumulate in the sub G1 phase: 2.19 % to 6.71 %. The cells in the G1 phase decrease from 77.3 % (day one) to 73.8 % (day three) but having the highest accumulation on

day two with 85.6 %. The S phase contains between 0.19 % and 1.56 % of the NY-BR-1 GFP$^+$ cells. The amount of cells in the G2 phase decreases from 20.3 % (day one) to 17.9 % (day three) but with the lowest peak of 12.1 % on day two (figure 4.31D).

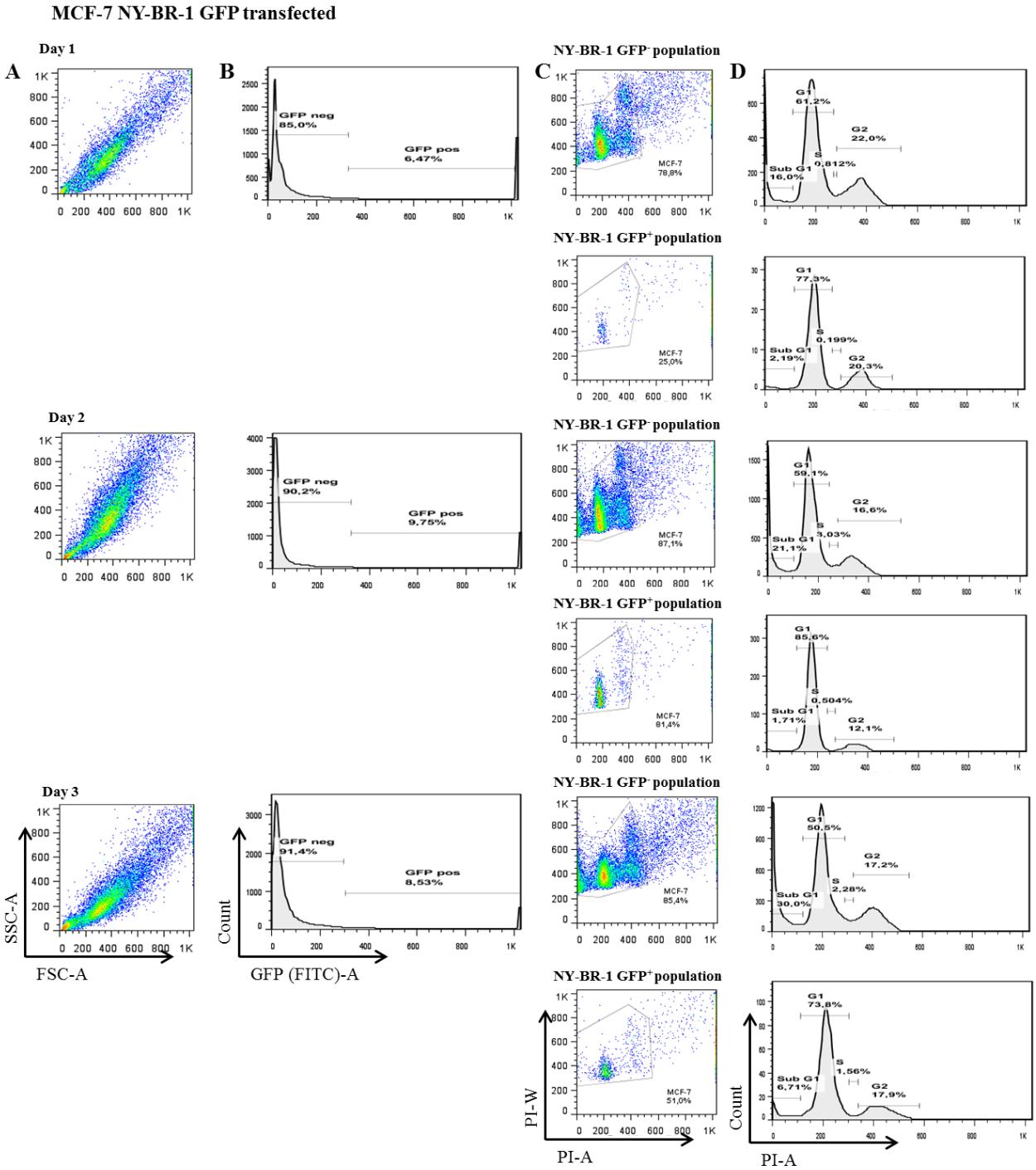


Figure 4.31: Results of the cell cycle analysis of MCF-7 NY-BR-1 GFP transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell debris out in either NY-BR-1 GFP negative or NY-BR-1 GFP positive cell population. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

4.3.3 Ki-67 staining in normal breast tissue

Serial FFPE tissue sections of healthy patients were generated to analyse the co-localization of NY-BR-1 with the proliferation marker Ki-67.

The Ki-67 staining (left) and the NY-BR-1 staining (right) of three different patients are shown exemplary in figure 4.32. The epithelial cells of patient HD-T-280 are Ki-67 negative but NY-BR-1 positive. Epithelial cells of the other two patients (HD-T-292, HD-T-311) are Ki-67 positive but NY-BR-1 negative.

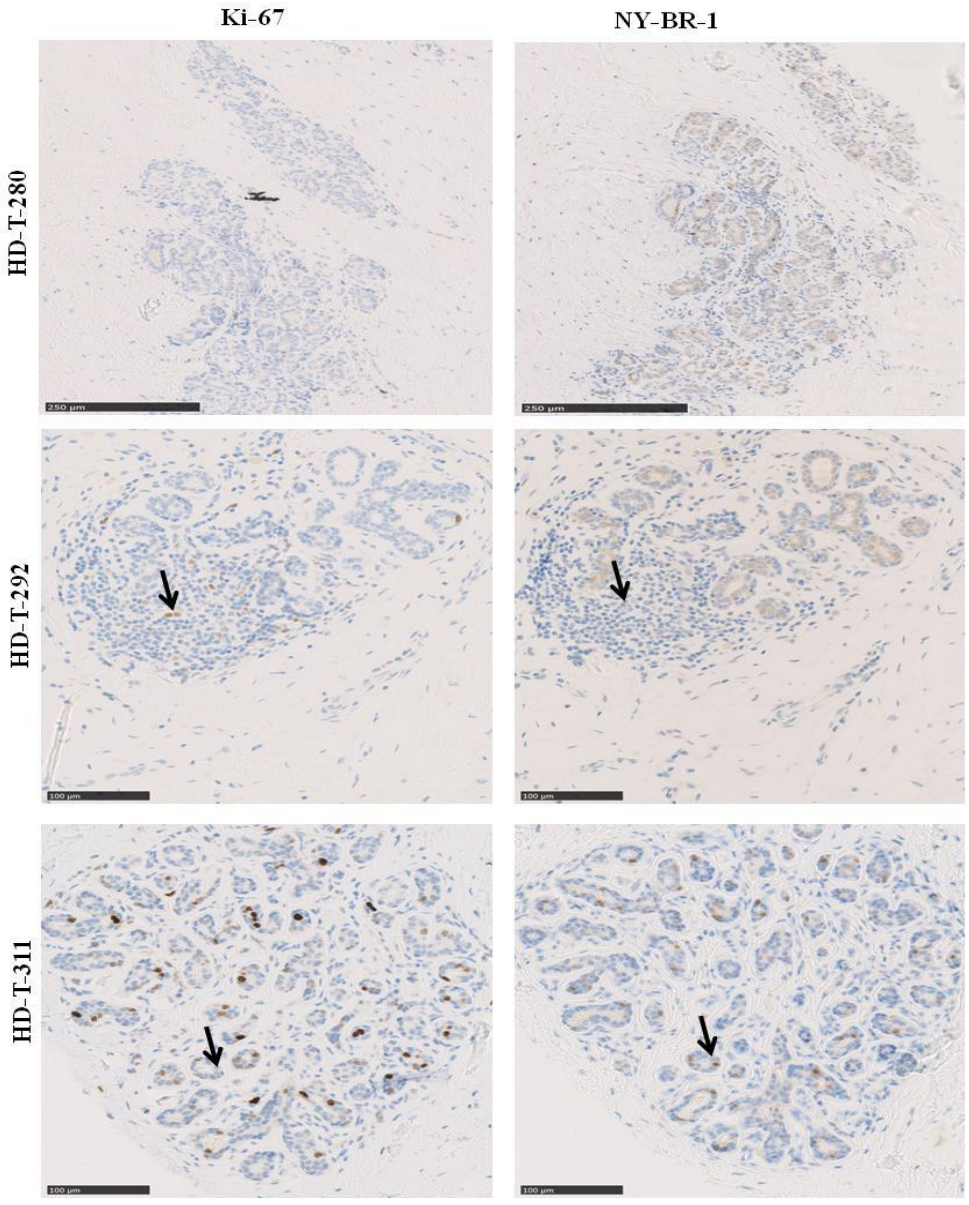


Figure 4.32: IHC staining of 4 µM FFPE tissue sections derived from three healthy patients. The nucleus is stained blue (haematoxylin) and the brown stained cells are Ki67 positive (left) and NY-BR-1 positive (right). The tissues derived from breast reductions. The magnification is 10 fold (HD-T-283) or 20 fold (HD-T-292/311).

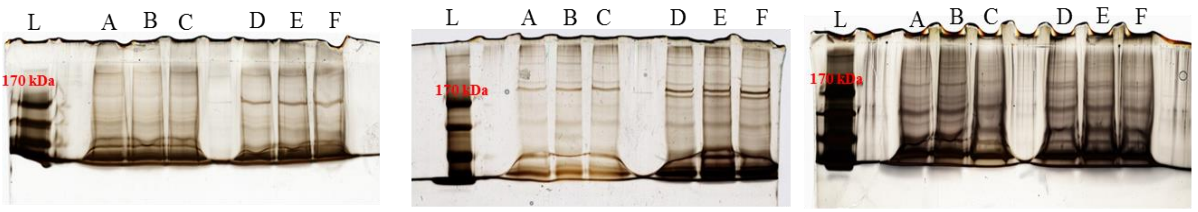
To summarize the results of this chapter: the proliferation assay showed in all four tested cell lines that NY-BR-1 GFP positive cells do not proliferate compared to the non-transfected and GFP transfected cell populations. Furthermore, the results of the cell cycle analysis indicate that NY-BR-1^{+} cells accumulate more in G1 phase compared to the NY-BR-1^{-} fraction and non-transfected cells. IHC staining with Ki-67 and NY-BR-1 of serial tissue sections demonstrated that Ki-67 positive cells are not NY-BR-1 positive.

4.4 Interaction partner of NY-BR-1

All proteins interact in a dynamic network to regulate cellular processes in a time-dependent manner. They have to communicate with each other to react to signals, stimulus or a specific cell state. The proteins form stable or transient complexes depending on their location and involvement of cellular processes and thus interact with each other.

Co-IP belongs to the strongest biochemical approaches to identify specific interaction partner of a protein of interest. Thereby an antibody, directed against the protein of interest, is coupled on beads containing immobilized protein A or protein G. This complex "fishes" then from cell lysates the protein-protein complex. The Co-IP is combined with mass spectrometry (MS) to reveal potential interacting partners. The advantages of this method are low-cost, ease-of-use, compatibility with other downstream methods and relatively high specificity (Ngounou Wetie et al., 2014).

To classify significant interaction partner of NY-BR-1 three cell lines (HEK293, MCF-10A, and MCF-7) were either transfected with the empty pc3 vector (as a control) or pc3 NY-BR-1. Co-IP reactions were performed in triplicates with the α-NY-BR-1 (C18) antibody. The IP lysates were loaded onto a SDS-Page gel, a silver staining was performed and the gels (figure 4.33) were brought to the proteomics core facility of the German Cancer Research Center for mass spectrometry analysis.



A-C: HEK293 transfected with pc3 vector
D-F: HEK293 transfected with pc3 NY-BR-1

A-C: MCF-10A transfected with pc3 vector
D-F: MCF-10A transfected with pc3 NY-BR-1

A-C: MCF-7 transfected with pc3 vector
D-F: MCF-7 transfected with pc3 NY-BR-1

Figure 4.33: SDS Page gels after mass spectrometry compatible silver staining for mass spectrometry analysis. The cell lines HEK293, MCF-10A and MCF-7 were transfected either with pc3 empty vector (lane A-C) or pc3 NY-BR-1 (lane D-F).

4.4.1 Mass spectrometry results

Tables 4.4 - 4.6 show the unique proteins that were identified in each NY-BR-1 expressing cell line (transiently transfected MCF-10A, MCF-7, and HEK293) compared to the respective control cells and thus represent possible interaction partners of NY-BR-1.

The identification of proteins from peptide sequences was done with the Mascot software. Before entering mass spectrometry analysis the proteins are digested to form smaller peptides, which can be analysed by a mass spectrometer. During analysis the molecular weight of the peptides is measured and then compared by Mascot with database known peptides. Considering the cleavage enzyme used for digestion Mascot can calculate a theoretical mass for each peptide. Next a score is calculated based on the probability that the peptides match those in the database. The evaluated score in the raw data describes how many peptides Mascot could identify from a particular protein: the more the higher the score (Koenig et al., 2008).

For our analysis the merge files were used. The lanes were sliced prior mass spectrometry analysis and thus a protein could be present in multiple slices. To receive an appropriate score the repeatedly present proteins were summated. The number of scores or significant protein sequences indicates that the protein was detected in an equivalent number of lanes. NY-BR-1 is displayed as AN30A and could be detected in all three cell lines in all three lanes. The differences between the NY-BR-1 scores/significant protein sequences between the three cell lines can be due to a difference in transfection efficacy resulting in different protein expression levels.

Table 4.4: Overview of the unique proteins detected in the IP fraction of pc3NY-BR-1 transfected MCF-10A cells.

MCF-10A NY-BR-1 (Unique)	Gene Name	Score	Sign Prot Sequences
ABCE1_HUMAN	ATP-binding cassette sub-family E member 1	115	2
ACTN4_HUMAN	Alpha-actinin-4	100	2
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	2593;1523;1765	42;29;32
ARF5_HUMAN	ADP-ribosylation factor 5	140	2
CH60_HUMAN	60 kDa heat shock protein, mitochondrial	83	2
GSTM2_HUMAN	Glutathione S-transferase Mu 2	130	2
GSTM3_HUMAN	Glutathione S-transferase Mu 3	122	2
HS71L_HUMAN	Heat shock 70 kDa protein 1-like	225;157	3;2
K22O_HUMAN	Keratin, type II cytoskeletal 2 oral	466;458	9;8
K2C7_HUMAN	Keratin, type II cytoskeletal 7	238;239	4;4
KV201_HUMAN	Ig kappa chain V-II region Cum	107	2
KV304_HUMAN	Ig kappa chain V-III region Ti	116	2
LDHB_HUMAN	L-lactate dehydrogenase B chain	96	2
PHB_HUMAN	Prohibitin	62	2

PRDX4_HUMAN	Peroxiredoxin-4	73	2
RAB2A_HUMAN	Ras-related protein Rab-2A	124	3
RL15_HUMAN	60S ribosomal protein L15	134;108	2;2
RL27A_HUMAN	60S ribosomal protein L27a	96	2
RL35A_HUMAN	60S ribosomal protein L35a	165	4
RL6_HUMAN	60S ribosomal protein L6	67	2
RS6_HUMAN	40S ribosomal protein S6	79	2
S10A9_HUMAN	Protein S100-A9	78	2
SMD3_HUMAN	Small nuclear ribonucleoprotein Sm D3	65	2
SPB3_HUMAN	Serpin B3	79	2
SPB7_HUMAN	Serpin B7	115	2
SRSF1_HUMAN	Serine/arginine-rich splicing factor 1	110;167	2;3
TBB4B_HUMAN	Tubulin beta-4B chain	477	8
TPM2_HUMAN	Tropomyosin beta chain	170	4
U520_HUMAN	U5 small nuclear ribonucleoprotein 200 kDa helicase	81;68	2;2

Table 4.5: Overview of the unique proteins detected in the IP fraction of pc3NY-BR-1 transfected MCF-7 cells.

MCF-7 NY-BR-1 (Unique)	Gene Name	Score	Sign Prot Sequences
ACTA_HUMAN	Actin, aortic smooth muscle	307	4
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	787;813;1091	13;16;21
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	135	3
EPIPL_HUMAN	Epiplakin	166	4
G6PD_HUMAN	Glucose-6-phosphate 1-dehydrogenase	139	2
HS71L_HUMAN	Heat shock 70 kDa protein 1-like	146;170	2;3
K22O_HUMAN	Keratin, type II cytoskeletal 2 oral	380	7
K2C75_HUMAN	Keratin, type II cytoskeletal 75	676	10
KRT34_HUMAN	Keratin, type I cuticular Ha4	649	11
KRT35_HUMAN	Keratin, type I cuticular Ha5	128	2
KRT81_HUMAN	Keratin, type II cuticular Hb1	591	9
KRT84_HUMAN	Keratin, type II cuticular Hb4	255	5
KT33B_HUMAN	Keratin, type I cuticular Ha3-II	769	13
MYH10_HUMAN	Myosin-10	155	3
MYH14_HUMAN	Myosin-14	109	2
PRDX2_HUMAN	Peroxiredoxin-2	90	2
RCN2_HUMAN	Reticulocalbin-2	223	3
RL12_HUMAN	60S ribosomal protein L12	74	2
RLA0_HUMAN	60S acidic ribosomal protein P0	103	2
RLA0L_HUMAN	60S acidic ribosomal protein P0-like	123	2
SNX27_HUMAN	Sorting nexin-27	162;182	3;4
SRSF3_HUMAN	Serine/arginine-rich splicing factor 3	86	2
SRSF7_HUMAN	Serine/arginine-rich splicing factor 7	177	3
TBB4B_HUMAN	Tubulin beta-4B chain	266;484;462	4;7;7

Table 4.6: Overview of the unique proteins detected in the IP fraction of pc3NY-BR-1 transfected HEK293 cells.

HEK293 NY-BR-1 (Unique)	Gene Name	Score	Sign Prot Sequences
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	2717;2799;2836	47;40;43
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	181	5
GIT1_HUMAN	ARF GTPase-activating protein GIT1	132	2
H12_HUMAN	Histone H1.2	99	2
H4_HUMAN	Histone H4	79	2
IGHG3_HUMAN	Ig gamma-3 chain C region	115	2
K1C15_HUMAN	Keratin, type I cytoskeletal 15	570	9
K1C24_HUMAN	Keratin, type I cytoskeletal 24	200	3
K2C75_HUMAN	Keratin, type II cytoskeletal 75	705;649	10;11
KDIS_HUMAN	Kinase D-interacting substrate of 220 kDa	550;355;329	10;6;7
KRT34_HUMAN	Keratin, type I cuticular Ha4	214;301	4;5
KRT36_HUMAN	Keratin, type I cuticular Ha6	187	3
KRT83_HUMAN	Keratin, type II cuticular Hb3	313	5
PLEC_HUMAN	Plectin	116	2
RL7A_HUMAN	60S ribosomal protein L7a	71	2
SATB1_HUMAN	DNA-binding protein SATB1	93;146	2;3
SNX27_HUMAN	Sorting nexin-27	1076;1096	19;18
TBA1A_HUMAN	Tubulin alpha-1A chain	349;430	4;7
TBB4B_HUMAN	Tubulin beta-4B chain	273	5
TR150_HUMAN	Thyroid hormone receptor-associated protein 3	147;185;103	2;3;2
VANG1_HUMAN	Vang-like protein 1	84;119	2;2
ZO2_HUMAN	Tight junction protein ZO-2	303;193;346	7;4;7
PRDX1_HUMAN	Peroxiredoxin-1	75	2

The unique proteins of the NY-BR-1 fraction of each cell line were compared to see whether some possible interaction partners of NY-BR-1 is shared between all three tested cell lines.
In MCF-7 and HEK293 six proteins, in MCF-10A and MCF-7 cells four proteins and in MCF-10A and HEK293 cells just two proteins are shared including NY-BR-1 (Table 4.7-4.9). The HS71L protein could be detected in two lanes in both cell lines (MCF-10A/MCF-7). This Heatshock protein is involved to maintain the right folding of a newly synthetized protein (chaperone character) and to prevent aggregation of existing proteins (http://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=3305). The other two proteins (K220, TBB4B) could be detected in two lanes in just one of the compared cell line (table 4.7). K220, also known as KRT76, belongs to the type II keratin family. This family includes keratins constitute for type II intermediate filaments of the intracytoplasmatic cytoskeleton of epithelial cells. Keratins help to maintain the normal tissue structure and function (Ambatipudi et al., 2013). KRT76 is mainly expressed in the suprabasal epithelial

cells of the hard palate and gingiva and is down-regulated in oral cavity tumours (Ambatipudi et al., 2012; Collin et al., 1992).

Table 4.7: Shared proteins between MCF-10A/MCF-7 cells in the NY-BR-1 fraction.

MCF-10A/MCF-7 (Shared)	Gene Name	Score MCF-10A	Sign Prot Sequences MCF-10A	Score MCF-7	Sign Prot Sequences MCF-7
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	2593;1523;1765	42;29;32	787;813;1091	13;16;21
HS71L_HUMAN	Heat shock 70 kDa protein 1-like	225;157	3;2	146;170	2;3
K22O_HUMAN	Keratin, type II cytoskeletal 2 oral	466;458	9;8	380	7
TBB4B_HUMAN	Tubulin beta-4B chain	477	8	266;484;462	4;7;7

The ATPA protein is present in both cell lines but in just one lane and this protein is involved in the respiratory chain. The two keratin proteins (K2C75, KRT34) are once detected in MCF-7 but twice in the HEK293 cells and are found in hairs and nails. The task of SNX27 is endocytosis of plasma membrane receptors and it supports their protein trafficking (Worby and Dixon, 2002). This protein is found in both cell lines in two lanes. TBB4B can be seen in all three lanes of MCF-7 but just in one of HEK293 (table 4.8).

Table 4.8: Shared proteins between MCF-7/HEK293 cells in the NY-BR-1 fraction.

MCF-7/HEK293 (Shared)	Gene Name	Score MCF-7	Sign Prot Sequences MCF-7	Score HEK293	Sign Prot Sequences HEK293
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	787;813;1091	13;16;21	2717;2799;2836	47;40;43
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	135	3	181	5
K2C75_HUMAN	Keratin, type II cytoskeletal 75	676	10	705;649	10;11
KRT34_HUMAN	Keratin, type I cuticular Ha4	649	11	214;301	4;5
SNX27_HUMAN	Sorting nexin-27	162;182	3;4	1076;1096	19;18
TBB4B_HUMAN	Tubulin beta-4B chain	266;484;462	4;7;7	273	5

MCF-10A and HEK293 have only TBB4B in common besides NY-BR-1. TBB4B could be detected just in one lane in both cell lines (table 4.9).

Table 4.9: Shared proteins between MCF-10A/HEK293 cells in the NY-BR-1 fraction.

MCF-10A/HEK293 (Shared)	Gene Name	Score MCF-10A	Sign Prot Sequences MCF-10A	Score HEK293	Sign Prot Sequences HEK293
AN30A_HUMAN	Ankyrin repeat domain-containing protein 30A	2593;1523;1765	42;29;32	2717;2799;2836	47;40;43
TBB4B_HUMAN	Tubulin beta-4B chain	477	8	273	5

TBB4B is the only protein shared by all three cell lines. This protein is the major component of microtubules. Microtubules form the cytoskeleton of each cell and are involved in many cellular processes such as chromosome separation during mitosis/meiosis and cellular trafficking (Vicente and Wordeman, 2015).

4.5 Is NY-BR-1 a progenitor cell marker?

Stem or progenitor cells are important to maintain tissues and to replace lost cells during normal attrition and injuries. They are present in every tissue. The mammary progenitor cells help to shape the mammary glands during puberty, pregnancy, and lactation. Different subsets of progenitor cells in the mammary gland are known as described in the introduction (see 1.4).

The proliferation assay showed that NY-BR-1 positive cells do not proliferate but also not going into an increased apoptosis compared to NY-BR-1 negative cells. The analysis of the cell cycle revealed that more NY-BR-1 positive cells accumulate in the G1 phase compared to the non-transfected and NY-BR-1 negative cell population. As described in chapter 4.1.1 the mosaic-like expression pattern of the NY-BR-1 protein suggests that the NY-BR-1 positive cells form a functionally different cell population in the mammary gland. The NY-BR-1 expression pattern in breast tumours depends on the molecular subtype (figure 4.44). If the theory is considered that the different subtypes in breast cancer are due to the different progenitor population it can be probable that NY-BR-1 is a marker for a certain progenitor cell population in the mammary gland.

4.5.1 Generation and analysis of mammospheres

Several years ago an *in vitro* cultivation system was developed to generate non-adherent mammospheres from human mammary epithelial cells. The characteristic feature of these cells is that they are undifferentiated due to the fact that they can proliferate in suspension. Additionally, it was shown that these mammospheres have the ability to differentiate along the mammary epithelial lineages (Dontu et al., 2003).

Thus, in theory only undifferentiated cells have the ability to form mammospheres in ultra-low binding plates. After seven to nine days of culturing, primary spheres can be harvested, after two weeks secondary and after three weeks tertiary. They are analysed via fluorescence staining and qPCR. In figure 4.34 primary, secondary and tertiary mammospheres of patient HD-T-311 are shown prior to analysis. The isolated epithelial cells were isolated from healthy breast tissue.

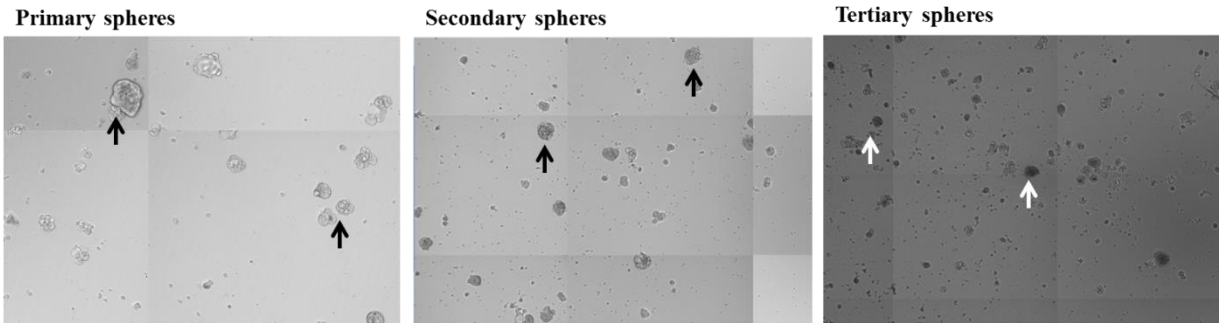


Figure 4.34: Primary, secondary and tertiary mammospheres of isolated healthy epithelial cells from patient HD-T-311. The arrows indicate spheres.

4.5.2 qPCR analysis of mammospheres

To analyse the expression of NY-BR-1 and other lineage markers, the spheres were harvested, RNA was isolated and a qPCR was performed. First, the NY-BR-1 expression in four patients was tested (figure 4.35A). The NY-BR-1 expression in the primary mammospheres of patients HD-T-263 (C_P: 25) and HD-T-311 (C_P: 28) is higher compared to the patients HD-T-283 (lowest expression profile) (C_P: 31) and HD-T-292 (C_P: 28). All samples were normalized against the housekeeping gene ß-Actin (C_P: 13-19).

It was possible to generate enough spheres from patient HD-T-311 to dissociate the primary ones to generate secondary and at the end even tertiary spheres. The results show that NY-BR-1 expression in the spheres is lost during longer culturing periods (figure 4.35B) with C_P values starting with 28 and decreasing to 35.

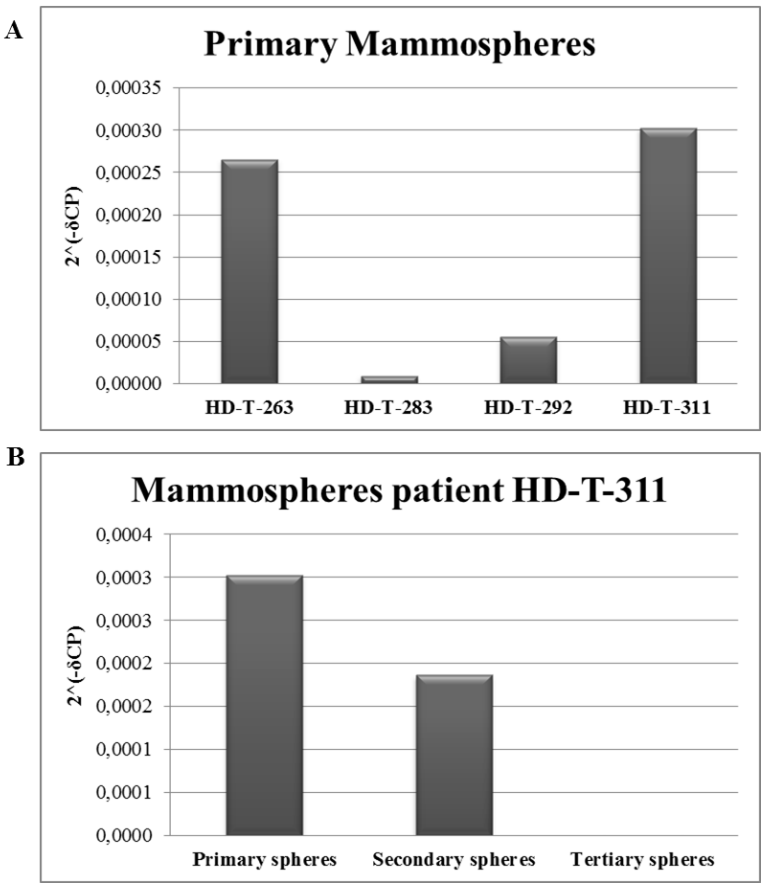


Figure 4.35: NY-BR-1 expression in mammospheres derived from isolated healthy epithelial cells assessed by qPCR. A: Primary mammospheres of four different patients were tested regarding their NY-BR-1 expression level. B: NY-BR-1 expression was analysed in primary, secondary and tertiary mammospheres of patient HD-T-311.

4.5.3 Fluorescence staining of primary mammospheres

Fluorescence staining was performed to test whether mammospheres also express NY-BR-1 on the protein level. Enough spheres were generated of two different healthy patients (figure 4.36). The spheres were stained with Hoechst (blue (nucleus), α-NY-BR-1#2 (green) and either with cytokeratin18 (CK18) or CK14 antibodies (red). A mammosphere is shown for patient HD-T-258 and three spheres for patient HD-T-311. In these two patients it can be observed that the NY-BR-1 expression is distributed on the membrane and in the cytosol of a cell. CK18 is a luminal cell cytokeratin and is used to discriminate myoepithelial from luminal cells. CK14 is a marker used to detect myoepithelial cells. No (HD-T-311) or just weak (HD-T-258) CK18 expression can be observed and this indicate that the cells are not luminal origin.

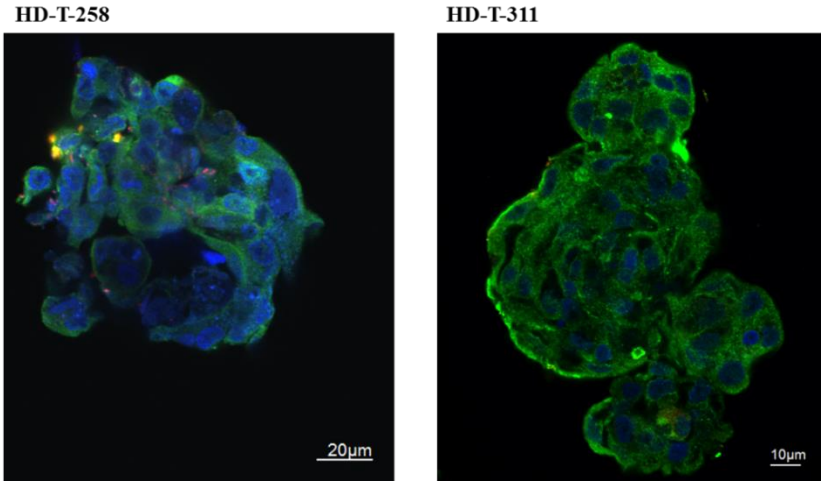


Figure 4.36: Fluorescence staining of mammospheres of two different patients. (Blue: nucleus, Green: NY-BR-1, Red: CK18 (HD-T-258), CK14 (HD-T-311))

The spheres were further analysed on RNA level to characterize the cell population in more detail because the different progenitor cell populations have different kinds of gene expression. Therefore, the expression level of HER2, PR and ER were investigated. All samples were normalized to the housekeeping gene ß-Actin. The expression levels of ER are not shown because in all tested samples a C_P value of 35 was detected. Thus, no ER expression on RNA level in all tested samples could be observed. The HER2 expression is in general higher (C_P: between 21 and 28) (figure 4.37A) and PR expression levels (C_P: 33 – 35) (figure 4.37B). HD-T-311 secondary spheres show the highest HER2 expression but do not express the PR. In the HD-T-292 spheres HER2 expression is the lowest and no PR expression could be detected at all.

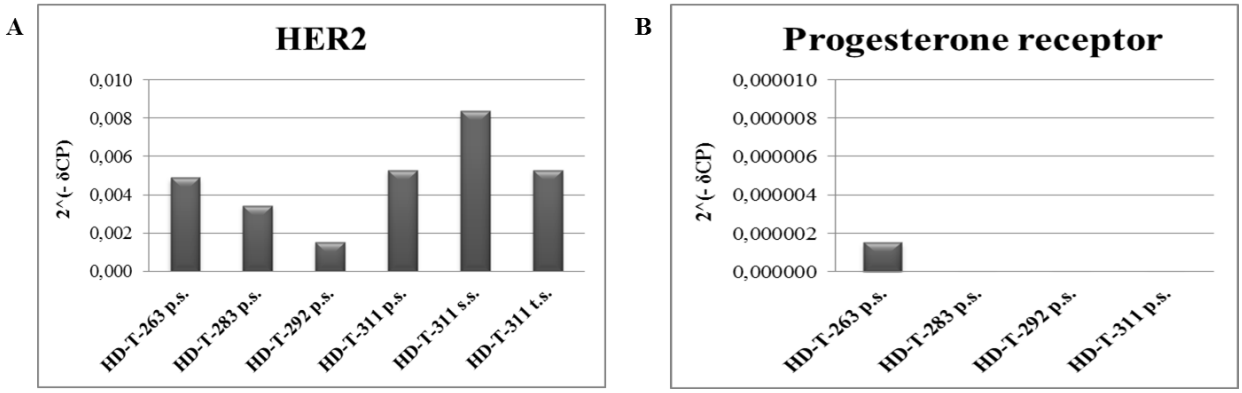


Figure 4.37: Expression level of Her2 (A) and progesterone receptor (B) in mammospheres of different patients analysed by qPCR. (p.s.; primary sphere, s.s.: secondary sphere)

The mammospheres were further characterized and the expression level of FOXA1 (also known as HNF-3A), GATA-3 and integrin-α 6 (also known as CD49f) were analysed (figure

4.38). GATA-3 is a transcription factor and it regulates the differentiation of the luminal cells in the mammary gland (Kouros-Mehr et al., 2006). Moreover, GATA-3 is involved in the ERα pathway and is considered to be a marker in ER^+ and ER^-/AR^+ cancer (Kouros-Mehr et al., 2008). GATA-3 interacts with FOXA1, which is a downstream target of GATA-3 in the mammary glands. FOXA1, also known as HNF-3α, is called a pioneer factor for ERα in $ERα^+$ breast cancer. These factors have the ability to bind directly on chromatin. Integrin-α 6 is an adhesion molecule and characterizes basal and bipopent progenitor cells in the mammary gland.

In all tested spheres, GATA-3 expression levels are higher compared to FOXA1 expression. In the secondary andin the tertiary spheres of patient HD-T-311 the GATA-3 expression increases. The lowest expression of GATA-3 was observed in patients HD-T-263 and HD-T-292 (figure 4.38A). The highest level of FOXA1 expression was detected in patient HD-T-311 primary spheres and it decreases from the primary to the tertiary spheres. The lowest expression of this pioneer factor is observed in the primary spheres of patient HD-T-283 and HD-T-292 (figure 4.38B). The expression level of integrin-α 6 are comparable between the tertiary spheres of patient HD-T-311, and primary spheres of HD-T-283 and HD-T-292. The lowest amount can be found in patient HD-T-263 and the highest in the secondary spheres of patient HD-T-311 (figure 4.37C).

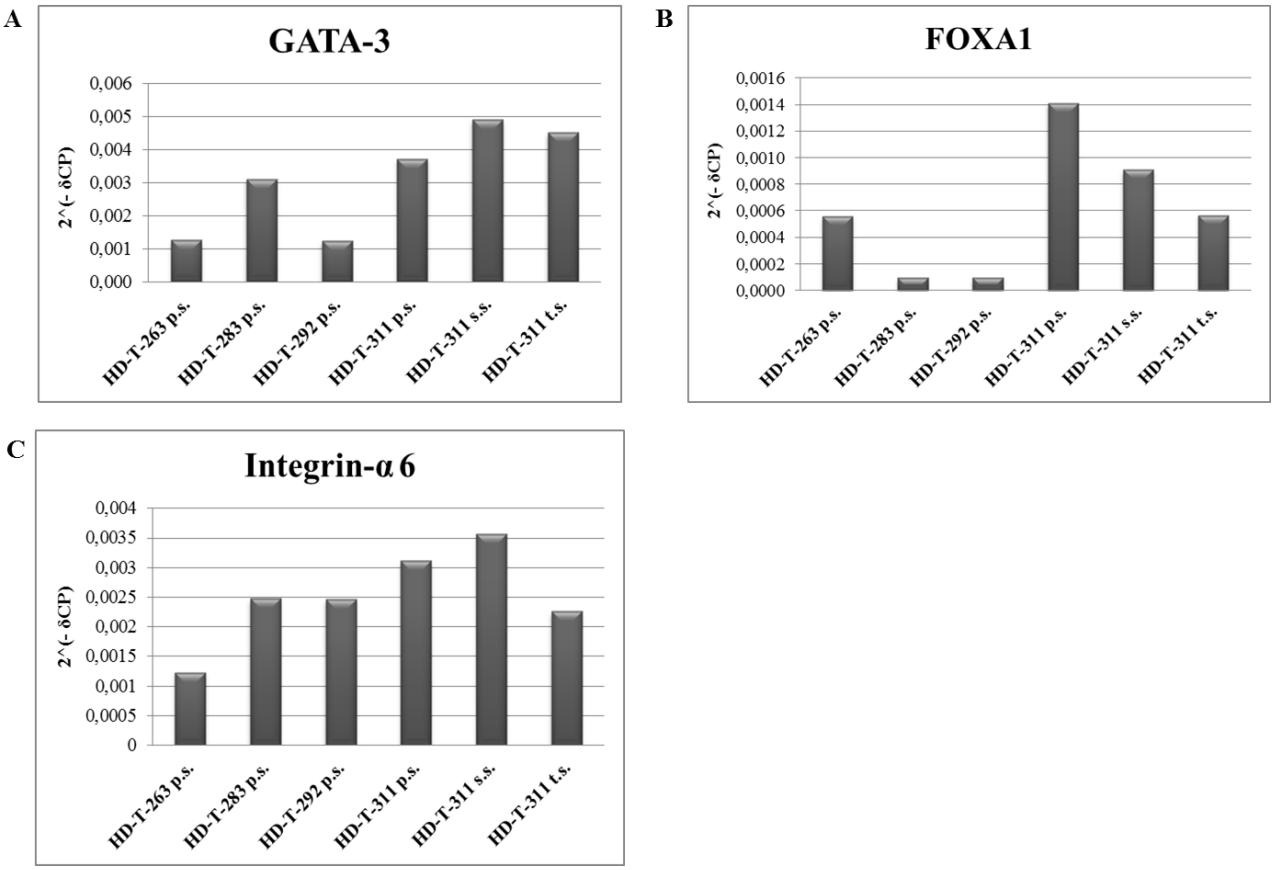


Figure 4.38: Expression profiles of GATA-3 (A), FOXA1 (B) and Integrin-α 6 (C) in the generated mammopsheres of all patients. (p.s.; primary sphere, s.s.: secondary sphere, t.s.: tertiary sphere)

4.5.4 IHC staining of ER and NY-BR-1 in serial tissue sections

Serial FFPE tissue sections of healthy patients were generated to analyse the co-localization of NY-BR-1 with ERα. The different progenitor subsets in the mammary gland can be ER^+ as well as ER^- as described in chaper 1.4.

The ERα (left) and the NY-BR-1 staining (right) of patient HD-T-292 and HD-T-311 are shown exemplary in figure 4.39. The mammary gland cells of patient HD-T-292 can be ER^- /NY-BR-1^+ (indicated with black arrows) as well as ER^+/NY-BR-1^+ (indicated with blue arrows). The cells of patient HD-T-311 consist of ER^-/NY-BR-1^+ cells.

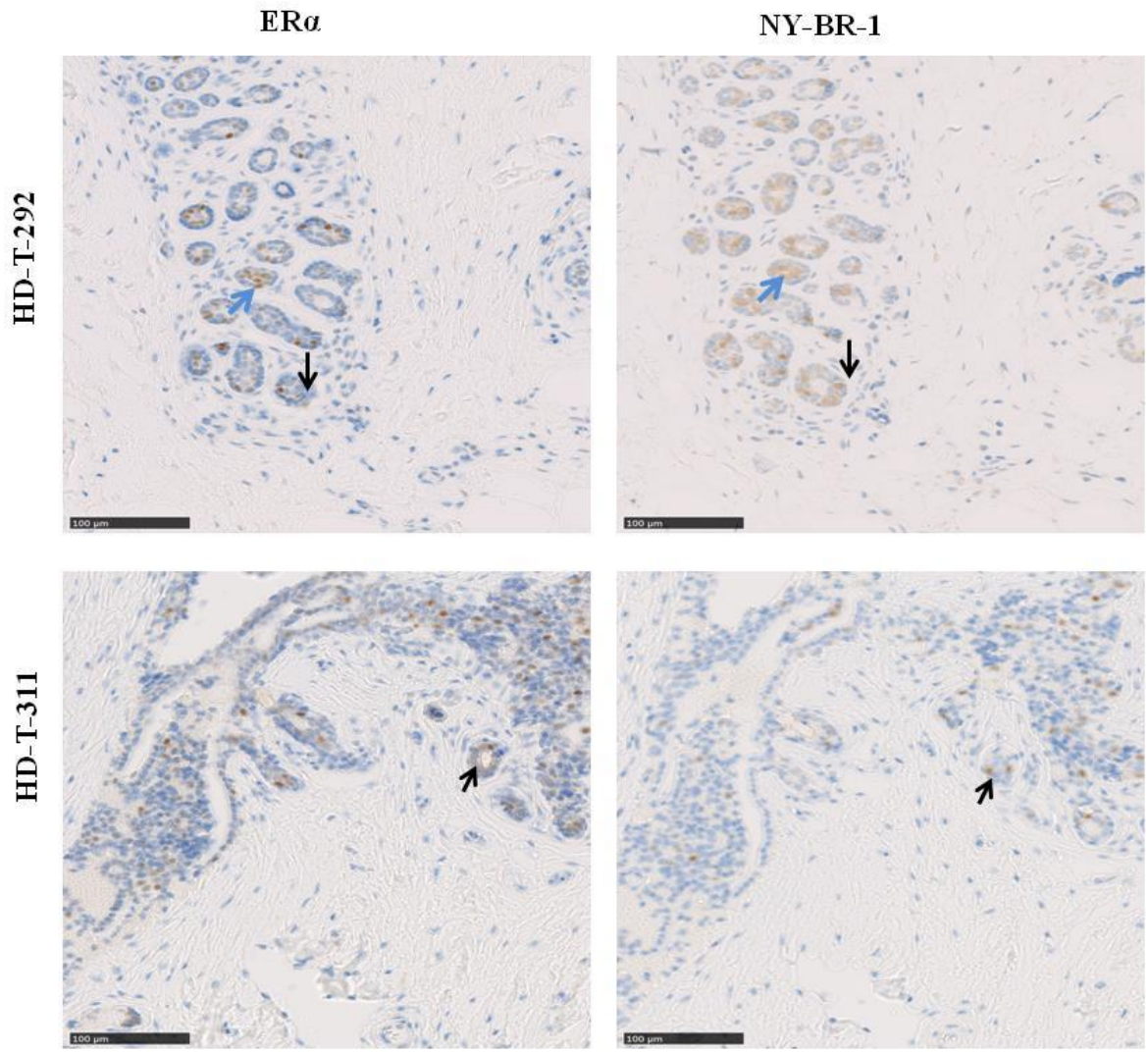


Figure 4.39: IHC staining of 4 µM FFPE tissue sections derived from healthy patients. The nucleus is stained blue (haematoxylin) and the brown stained cells are ER (left) and NY-BR-1 positive (right). The tissue derived from breast reduction. The magnification is 20 fold. The blue arrows indicate $ER^+/NY\text{-}BR\text{-}1^+$ cells, the black arrows $ER^-/NY\text{-}BR\text{-}1^+$.

In summary, a mosaic like NY-BR-1 expression pattern is observed in healthy patients. In breast carcinoma patient the expression pattern depends most likely on the molecular subtype ranging from no to a homogenous expression pattern. NY-BR-1 expression can be detected on RNA level for a specific time period in mammospheres but it becomes less after two weeks of culturing. Moreover, the expression level varies from patient to patient also depending on the progenitor subset. Co-expression with integrin-α6, HER2, GATA-3 and FOXA1 could be observed but not with ER and PR. The IHC staining revealed that NY-BR-1 positive cells are not always ER or PR positive.

4.6 *In silico* analysis

Integrative functional analyses involving genomics, transcriptomics and epigenomics were performed to explore the transcriptional regulation and expression pattern of NY-BR-1. For *in silico* analyses, data from publicly available resources such as dbSNP, ICGC, TCGA, ENCODE, and NCBI SRA were used. More information about the SNP analysis within the NY-BR-1 gene is depicted in the manuscript (page 168).

4.6.1 SNP analysis of the promoter region and gene body

SNPs in the promoter region can alter transcription factor binding and thus prevent or enhance transcription initiation. The SNPs, located in the promoter region of NY-BR-1 up to 2000 bp upstream from the transcription start site, are summarized in table 4.10. Additionally, the transcription factors are shown, predicted by the Alibaba 2.0 tool. Interestingly, SNPs exist in the predicted binding sites of GR, TBP, SP1, HNF-3, C/EBPα, YY1, c-Fos, c-Jun and AP-1. These factors are either involved in the ERα pathway or in transcription initiation or are known as proto-oncogenes.

Table 4.10: Overview of all SNPs, located in the promoter region of NY-BR-1. Their position is upstream of the transcription start site. Additionally, the predicted transcription factors are shown for specific SNP positions.

Variation Name	Position	Position in promotor upstream	Variant Alleles	Predicted transcription factors
rs184032975	37412790	1995	C/T	
rs12221253	37412799	1986	C/T	
rs370025586	37412814	1971	G/A	
rs77426826	37412816	1969	A/G	
rs189011172	37412850	1935	C/T	
rs371244298	37412860	1925	C/T	
rs368433595	37412889	1896	G/A	
rs374206891	37412903	1882	A/C	GR
rs371968048	37412937	1848	A/G	
rs113089428	37412967	1818	-/A	TBP, GATA-1, HB
rs191763434	37412968	1817	A/T	TBP, GATA-1, HB
rs199569240	37412968	1817	-/T	TBP, GATA-1, HB
rs182715277	37413056	1729	G/A	SP1
rs187045318	37413113	1672	G/A	
rs58721264	37413117	1668	C/T	SP1
rs377674377	37413121	1664	G/T	SP1
rs12357251	37413138	1647	A/G	HNF-3
rs7908887	37413139	1646	G/A	HNF-3
rs57894012	37413139	1646	-/AAAAAAAAAAA	HNF-3
rs12357252	37413144	1641	A/G	

rs191538949	37413189	1596	T/C	
rs11011043	37413218	1567	G/A	SP1
rs183689800	37413302	1483	C/T	SP1
rs187716252	37413346	1439	G/T	SP1, C/EBPα, AP-2α, E1, MyoD
rs11011044	37413352	1433	G/A	E1, MyoD
rs2944452	37413403	1382	T/C	
rs74772832	37413441	1344	C/A	SP1
rs71297788	37413453	1332	-/A	
rs192694302	37413498	1287	T/A	
rs150975043	37413524	1261	C/A	GR
rs115317596	37413572	1213	A/C	
rs147929641	37413593	1192	G/-	SP1, WT1, Krox-20
rs369399202	37413677	1108	G/A	SP1
rs117975813	37413682	1103	C/T	SP1, E4, myogenin
rs185741129	37413706	1079	G/C/T	
rs190448522	37413715	1070	C/T	SP1
rs145899299	37413763	1022	G/T	SP1
rs370828102	37413885	900	G/-	C/EBPα, HNF-3, TBP, HOXA4
rs78349175	37413887	898	A/T	C/EBPα, HNF-3, TBP, HOXA4
rs375624717	37413888	897	-/T	C/EBPα, HNF-3, TBP, HOXA4
rs74669586	37413889	896	T/A	C/EBPα, HNF-3, TBP, HOXA4
rs71007621	37413907	878	T/-	MEB-1, HNF-3
rs181762978	37413912	873	G/A	
rs200897330	37414007	778	TT/-	
rs147795150	37414033	752	G/A	
rs114832447	37414163	622	C/A	SP1, NF-1
rs35660139	37414241	544	G/A	Krox-20, RAP1, SP1, YY1
rs184411029	37414312	473	G/C	SP1
rs11011045	37414433	352	G/A	c-Jun, c-Fos, C/EBPα, AP-1
rs370351587	37414541	244	G/A	
rs376482133	37414558	227	C/G	SP1
rs141172730	37414582	203	A/G	
rs115377424	37414625	160	G/A	
rs61866716	37414661	124	T/C	SP1
rs61866717	37414679	106	G/C	ETF
rs371009767	37414774	11	C/T	SP1

Figure 4.40 summarizes the analysis of the spectrum of SNPs on the nucleotide level in the promoter region of NY-BR-1. It shows a conserved profile with A>T/T>A transitions being the most frequent changes followed by A>G/T>C. Several insertions and deletions are occurring. Less frequently the G>A/C>T and G>C/C>G transitions can be observed.

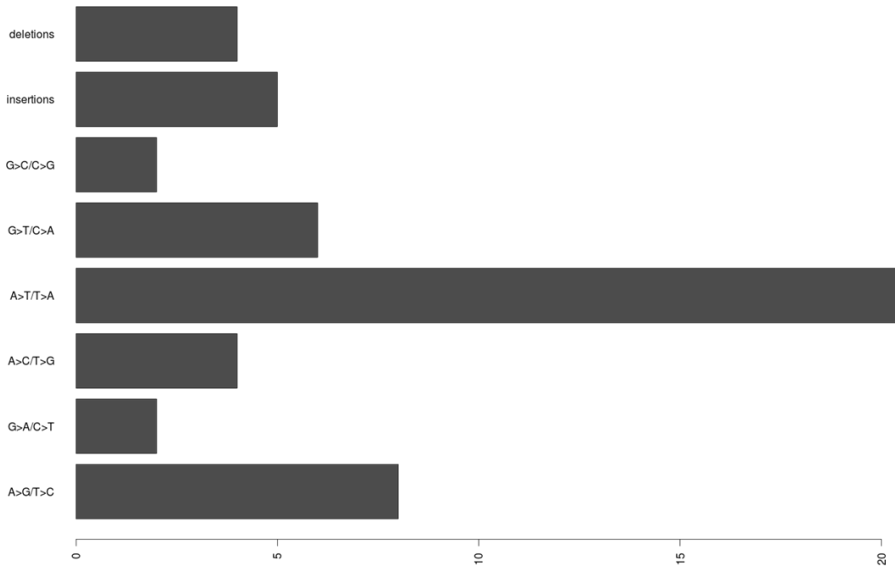


Figure 4.40: Graphical representation of the spectrum of nucleotide variations in the NY-BR-1 promoter.

Not only SNPs in a promoter region are important but also SNPs in the gene body because they can influence the splicing machinery. Thus, they might be responsible for new transcripts and proteins, which can have different structures and functions.

In the Ensembl BioMart database, 2898 SNPs are recorded for the ANKRD30A transcript ENST00000611781. 2880 of these were imported from dbSNP and 18 from NHLBI ESP. 1832 SNPs have been validated by independent submissions or frequency/genotype data. However, the clinical significance has not been determined yet for any of the SNPs. Out of all 2898 SNPs, 65 (2.07 %) were synonymous (sSNPs,) 191 (6.09 %) were non-synonymous (nsSNPs), and 2430 (77.48 %) occurred in intronic regions (figure 4.41A). nsSNPs were selected for further investigation. 69 nsSNPs were predicted to be damaging/deleterious by at least two of the used tools (PROVEAN, SIFT or PolyPhen). Analysis of the spectrum of nsSNPs on the nucleotide level showed a conserved profile with A>T/T>A transitions being the most frequent changes (figure 4.41B) The spectrum of amino acid changes was analysed, too. It could be observed that hydrophile>hydrophile transitions occur most frequently and hydrophobe>hydrophobe transitions second frequently (figure 4.41C). The nsSNPs rs200639888 and rs367841401, which were predicted to be highly damaging by all three tools, have an amino acid change from leucine to proline which are hydrophobe amino acids. The third damaging nsSNP predicted by all three tools, rs377750885, has a change from glutamic acid (hydrophile) to (hydrophobe) valine.

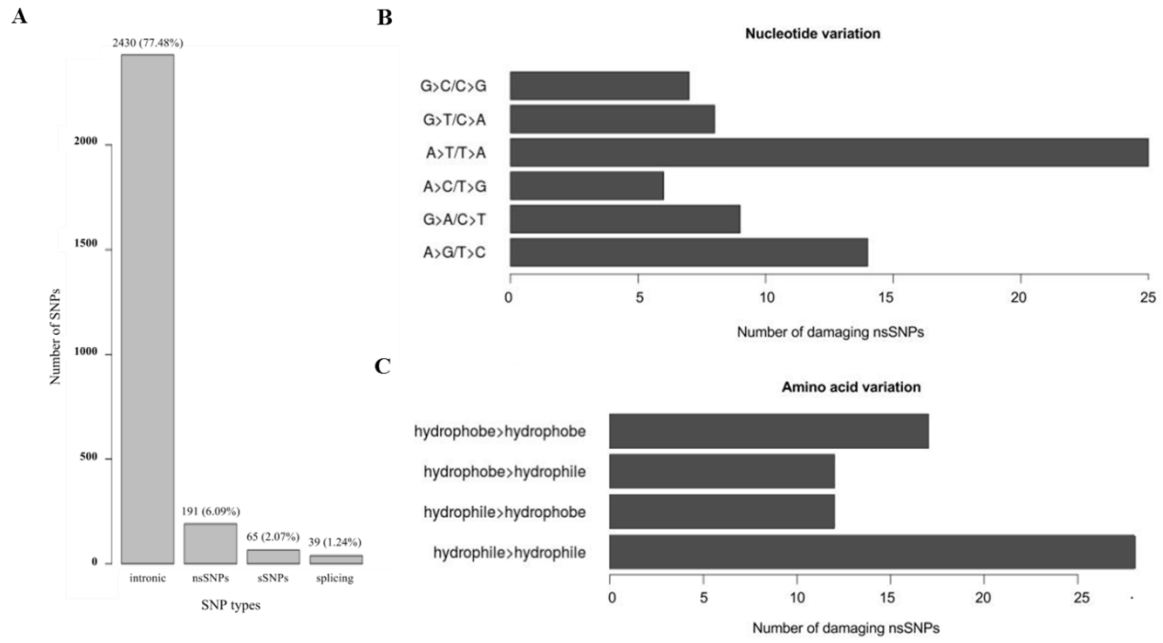


Figure 4.41: A: Graphical representation of distribution of intronic SNPs, non-synoymous SNPs (nsSNPs), synonymous SNPs (sSNPs), and SNPs at splicing sites for the NY-BR-1 gene, based on the dbSNP and NHLBI ESP databases. B+C: Graphical representation of spectrum of damaging nsSNPs variation. B) nucleotide variations, C) amino acid variations.

191 nsSNPs were analysed and the results varied between the used tools: SIFT predicted 79 damaging nsSNPs, Provean 28 nsSNPs, and PolyPhen2 102 nsSNPs. 16 nsSNPs were predicted to be damaging by all three tools, and a total of 69 nsSNPs were predicted damaging by at least two of the used tools. SIFT and PolyPhen2 have the biggest overlap with 63 common predictions. This may be due to the common step of assessing the degree of conservation by utilizing a multiple sequence alignment of homologous sequences. 36 damaging nsSNPs were only predicted by PolyPhen2 because PolyPhen2 is the only tool that takes functional relevant sites into account. The location of the 69 damaging SNP within the ANKRD30A gene is shown in figure 4.42A. It is noteworthy that a lot of the damaging SNPs are located in the longest exons 7 and 34. Between exon 19 and 24 no damaging nsSNPs are found and this can be observed in exon 35 and 36 as well. In dbSNP, 39 of the NY-BR-1 SNPs are annotated to be in splicing sites, located at donor or acceptor sites (figure 4.42B). These SNPs have the potential to influence the splicing process and new transcripts can be generated. The predicted splicing SNPs do not accumulate in exon 7 or 34 but more around exon 8, between exon 13 and 18, and between exon 25 and 31. A characteristic feature between the exons 10 and 30 is that the exons have a recurrent bp length (29, 73, 62, and 91) and contain repetitive elements.

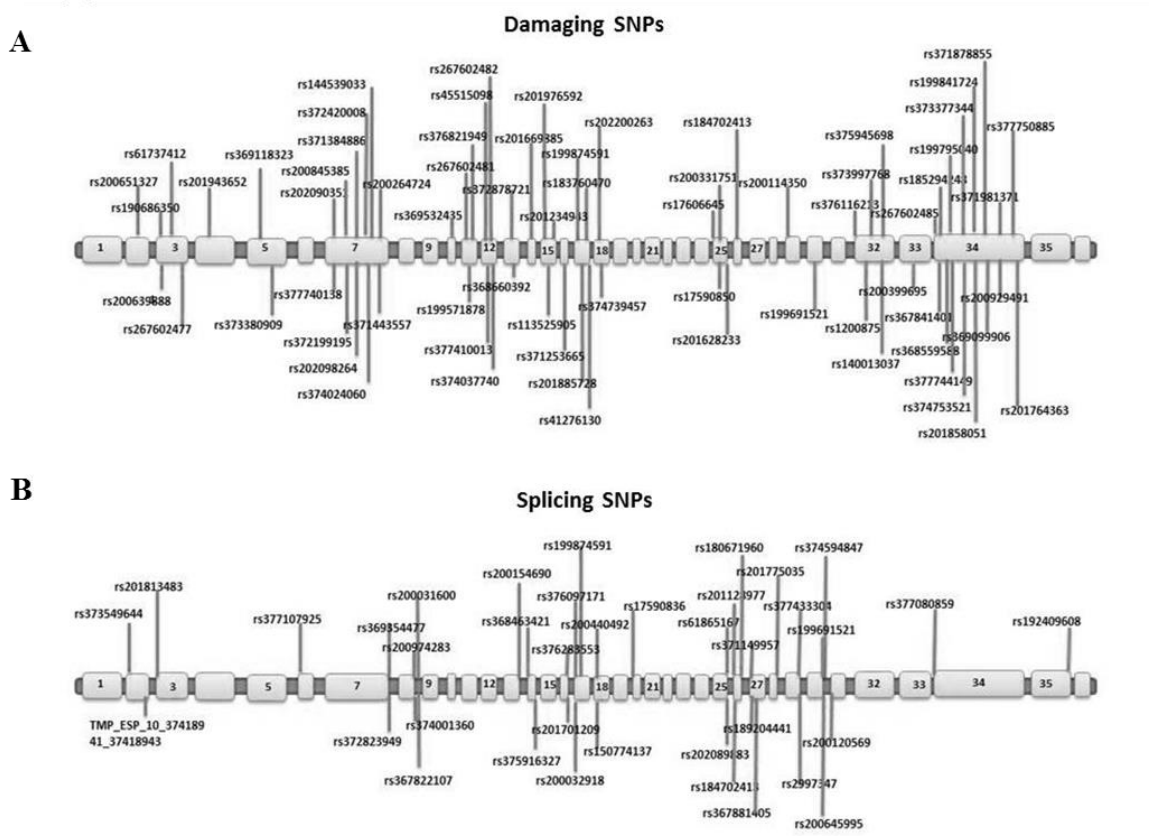


Figure 4.42: Location of the damaging SNPs (A) and splicing SNPs (B) within the gene of NY-BR-1 consisting of 36 exons. The length of exons is not true to scale.

4.6.2 NY-BR-1 expression in molecular subtypes of breast cancer

The molecular subtype classification is important in breast cancer to choose the best therapy strategy, which can differ significantly between hormone sensitive or insensitive tumours. This division depends on the receptor status ($ER^{+/-}$, $PR^{+/-}$, $HER2^{+/-}$). By using the publicly available RNA-Seq dataset published by Varley *et al.* which includes 28 samples of breast cancer cell lines, 42 samples of triple negative breast cancer (TNBC), 42 samples of ER^{+} breast cancer and 56 non-malignant breast tissue samples, it was possible to perform several analyses (Varley et al., 2014).

The first analysis included the NY-BR-1 gene expression in all samples. In few of the breast cell lines (BT-474, MDA-MB-134, DY30T2, MDA-MB-361, ZR-75-30, MDA-MB-453) a weak NY-BR-1 gene expression was detected. The highest gene expression can be observed in ER^{+} tumour samples but in all other samples (triple negative, uninvolved ER^{+}, healthy and uninvolved triple negative) a weak or no gene expression of NY-BR-1 can be seen (figure 4.43).

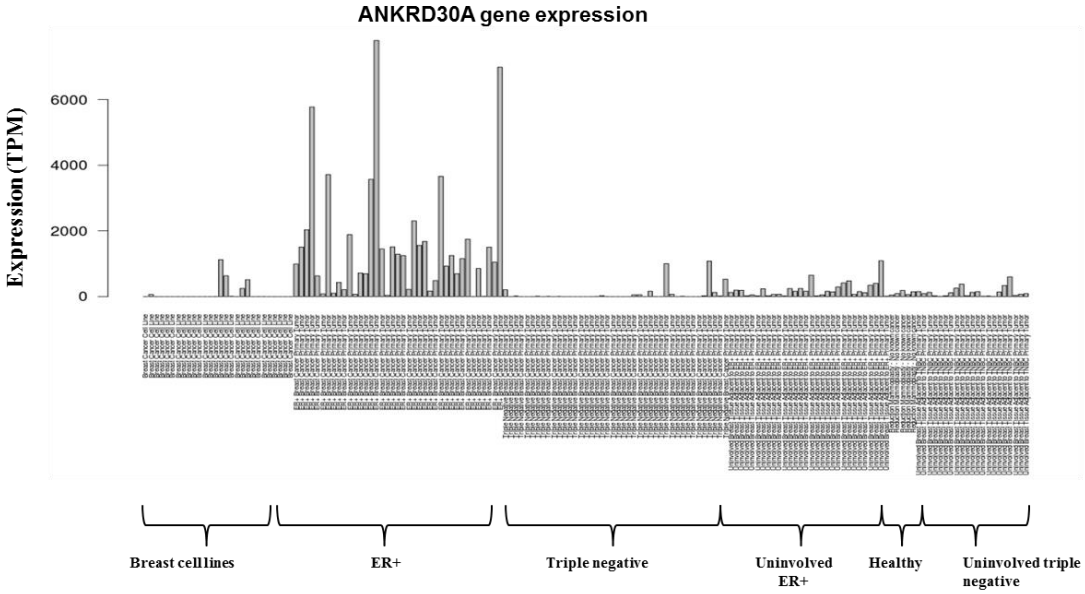


Figure 4.43: Gene expression analysis of NY-BR-1 in breast cell lines, ER^+ and triple negative tumors, uninvolved tissue of ER^+ and triple negative tumors and in healthy samples. This data set was published from Varley et al. (2014).

Within this data set the NY-BR-1 expression was compared in two contrary breast cancer subtypes (ER^+ and TNBC). Also the adjacent tissue was analysed. The results in figure 4.44 show that NY-BR-1 is significantly over-expressed in ER^+ compared to TN breast cancer samples. Moreover, NY-BR-1 is significantly over-expressed in ER^+ breast cancer (p=0.00003341) compared to the ER^+ adjacent healthy tissue. Between TNBC and its uninvolved tissue no significant difference in NY-BR-1 expression can be observed.

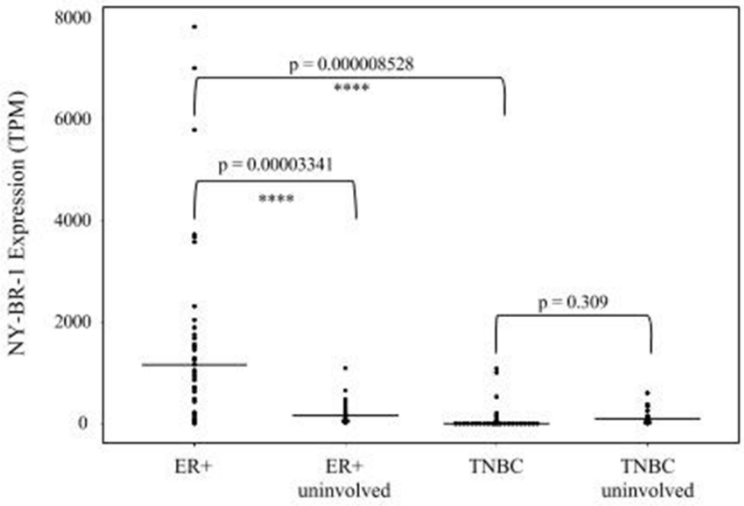


Figure 4.44: NY-BR-1 expression analysis in breast cancer subtypes and corresponding uninvolved tissue using the public available Varley data set. (TNBC: triple negative breast cancer)

Additionally an exon coverage analysis was performed. As mentioned before NY-BR-1 consists of 36 exons and it is likely that with such a high number of exons different transcripts might exist. Therefore, an exon coverage analysis was done with the Varley data set. Figure 4.45 shows that the first exons (one-nine) are rarely expressed in ER^+ breast cancer. The latter exons (26-36) have a higher expression profile in ER^+ breast cancer. In all other samples the exon coverage is in general very weak but it is striking that in ER^+ and TN uninvolved samples the latter exons are mainly expressed to a low extent. Thus, a part of NY-BR-1 is expressed on RNA level even in healthy tissue. In TN breast cancer samples no exon is highly expressed.

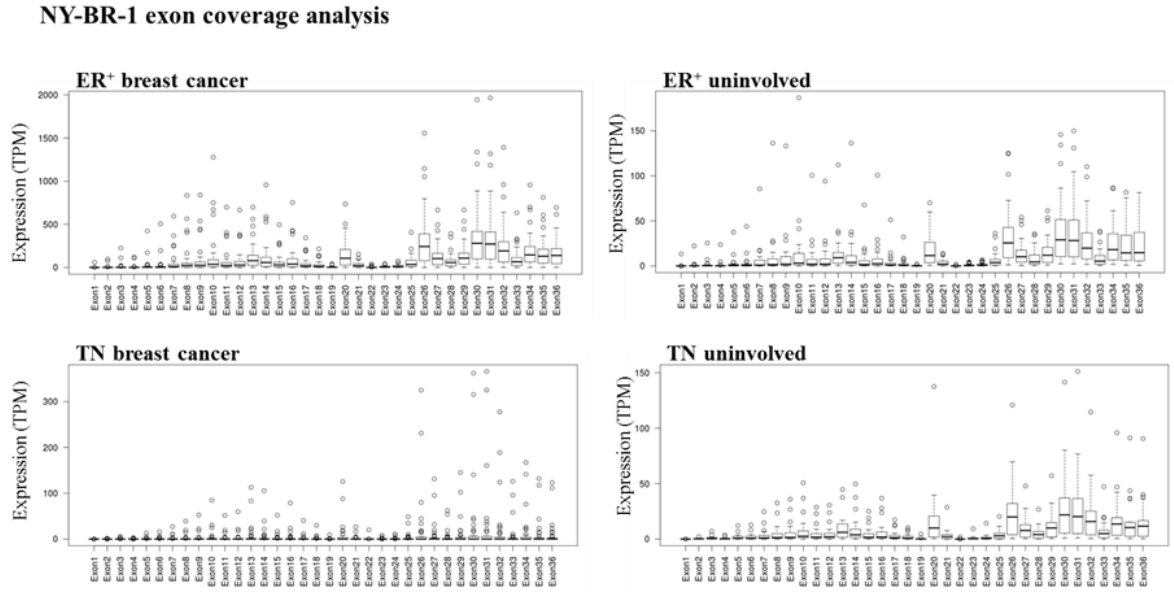


Figure 4.45: Exon coverage analysis of NY-BR-1 in ER^+ and TN breast cancer, ER^+ and TN uninvolved tissue samples.

Another analysis of the Varley et al. data set includes the genes, which are down-regulated in TNBC compared to ER^+ cancers. The more negative the fold change the more down-regulation of a gene takes place. The strongest down-regulated genes are on top and in table 4.11 the 50 most down-regulated genes in TNBC are shown. Interestingly, NY-BR-1 (rank 30) is more down-regulated than ESR1 (rank 37) but less than PGR (rank 24).

Table 4.11: Overview of 50 most down-regulated genes in TNBC of the Varley et al. data set.

Rank	Gene		Mean ER+	Mean TN	log2FoldChange
1	CLEC3A	C-type lectin domain family 3, member A	1.308378e+03	8.1556323	-7.325767
2	SERPINA6	serpin peptidase inhibitor, clade A, member 6	1.499204e+03	10.7746594	-7.120410
3	DSCAM-AS1	DSCAM antisense RNA 1	1.664696e+03	12.0825480	-7.106190
4	CDC20B	cell division cycle 20B	4.160569e+02	3.5834650	-6.859282

5	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1	7.126869e+03	79.3743597	-6.488452
6	GPR139	G protein-coupled receptor 139	1.386844e+01	0.1685781	-6.362244
7	CST9	cystatin 9 (testatin)	2.438800e+02	3.1042927	-6.295763
8	IL20	interleukin 20	4.106654e+02	5.9158212	-6.117241
9	CST5	cystatin D	9.000948e+01	1.5902817	-5.822723
10	POTEKP	POTE ankyrin domain family, member K, pseudogene	1.674666e+03	35.6991742	-5.551839
11	LINC01016	long intergenic non-protein coding RNA 1016	3.637712e+02	8.3872786	-5.438685
12	AGR3	anterior gradient 3	3.861895e+03	96.6420863	-5.320514
13	ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	5.678362e+03	156.1320543	-5.184636
14	TFF1	trefoil factor 1	4.883072e+03	135.5756561	-5.170619
15	TPM3_1	TPM3_1	6.820731e+01	1.9560138	-5.123938
16	KCNC2	potassium voltage-gated channel, Shaw-related subfamily, member 2	5.952804e+02	17.4759734	-5.090125
17	GRPR	gastrin-releasing peptide receptor	6.604642e+02	19.4249135	-5.087500
18	DEFB132	defensin, beta 132	5.315296e+01	1.6424929	-5.016191
19	CXorf27	chromosome X open reading frame 27	1.732960e+01	0.5483333	-4.982042
20	C1orf64	chromosome 1 open reading frame 64	1.148827e+03	38.1890170	-4.910859
21	KCNU1	potassium channel, subfamily U, member 1	2.547165e+01	0.8498496	-4.905541
22	TFF3	trefoil factor 3 (intestinal)	8.619518e+03	289.2790123	-4.897074
23	RIMS4	regulating synaptic membrane exocytosis 4	2.958008e+03	103.6255306	-4.835175
24	PGR	progesterone receptor	5.767404e+03	234.8600123	-4.618049
25	TPRG1-AS2	TPRG1 antisense RNA 2	3.033012e+02	12.4025144	-4.612047
26	FAM47A	family with sequence similarity 47, member A	1.122428e+01	0.4594130	-4.610687
27	AREG_1	AREG_1	1.031969e+02	4.4602995	-4.532116
28	NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	5.289695e+03	229.1042127	-4.529109
29	EEF1A2	eukaryotic translation elongation factor 1 alpha 2	1.759765e+04	792.7481463	-4.472377
30	ANKRD30A	ankyrin repeat domain 30A	1.456245e+05	7024.3568743	-4.373743
31	MIR548F2	microRNA 548f-2	9.005343e+00	0.4397379	-4.356065
32	CT62	cancer/testis antigen 62	3.205531e+02	15.7743010	-4.344915
33	LINC00993	LINC00993	1.183818e+04	586.0805833	-4.336204
34	GRIK3	glutamate receptor, ionotropic, kainate 3	4.330880e+03	219.3444329	-4.303390
35	LOC100131320	LOC100131320	5.768424e+02	29.4592284	-4.291386
36	NKAIN1	Na+/K+ transporting ATPase interacting 1	3.117204e+03	159.8424201	-4.285530
37	ESR1	estrogen receptor 1	1.347108e+04	695.0266169	-4.276654
38	C10orf82	chromosome 10 open reading frame 82	2.039416e+02	11.0623994	-4.204420
39	KLHDC7A	kelch domain containing 7A	1.411618e+03	76.8824712	-4.198551
40	CYP4B1	cytochrome P450, family 4, subfamily B, polypeptide 1	3.321585e+03	182.1837529	-4.188406
41	KCNK15	potassium channel, subfamily K, member 15	2.669967e+03	148.0443570	-4.172721
42	TPRG1	tumor protein p63 regulated 1	1.644171e+03	91.7290773	-4.163838
43	MAGEB17	melanoma antigen family B, 17	1.555129e+02	8.6969135	-4.160387
44	SPPL2C	signal peptide peptidase like 2C	5.785534e+01	3.2471110	-4.155222
45	CST9L	cystatin 9-like	1.520535e+01	0.8700051	-4.127411
46	CLIC6	chloride intracellular channel 6	5.775863e+03	332.9907089	-4.116483
47	LOC389033	placenta-specific 9 pseudogene	1.156680e+02	6.6923381	-4.111336
48	SYTL5	synaptotagmin-like 5	7.027512e+02	40.9525252	-4.100990
49	TNRC18P1	TNRC18P1	1.163228e+03	68.2845151	-4.090431
50	SYT9	synaptotagmin IX	5.237676e+02	30.8791238	-4.084223

4.6.3 Correlation of NY-BR-1 expression with other genes

To get a hint, which genes are up- or down-regulated while NY-BR-1 is expressed or not, an analysis of different data sets was performed. Analysis of the breast cancer ER ChIP-Seq dataset from Ross-Innes *et al.* revealed several ER binding sites in the genomic region of NY-BR-1 (figure 4.46) (Ross-Innes et al., 2012). It is striking that these binding sites are present from the beginning of the gene until intron 21. No binding sites are present from exon 23 until the end of the NY-BR-1 gene. The ER binding site between exon eight and nine exists in all three samples.

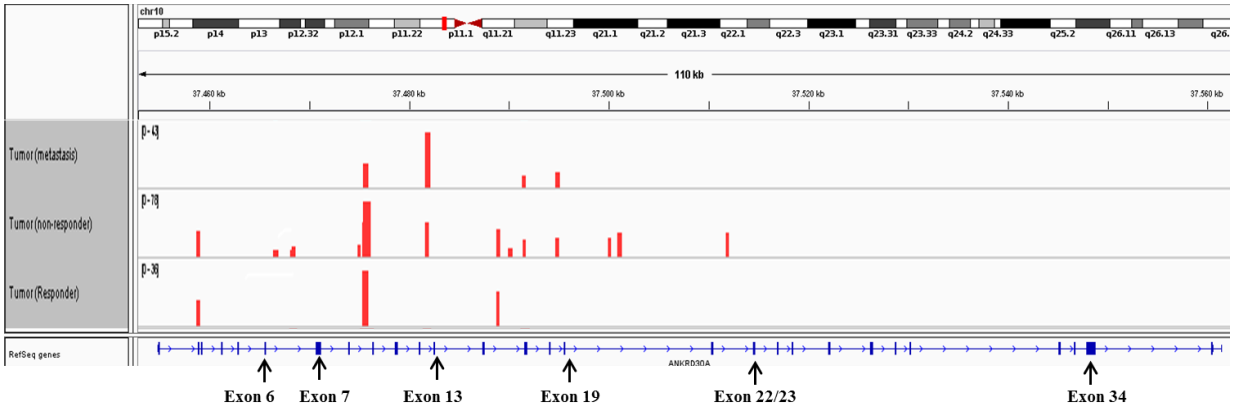
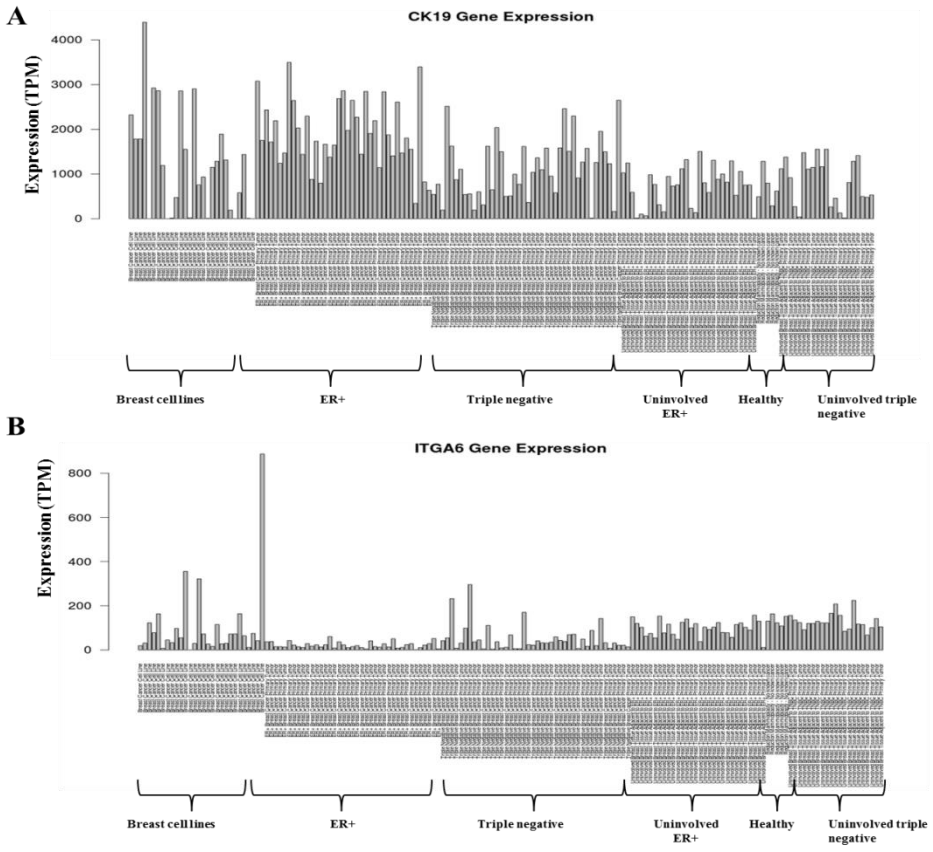


Figure 4.46: NY-BR-1 ER-binding peaks in three tumor samples detected with ER ChIP-seq.

Using the RNA-seq data from Varley et al. a gene expression profiling of the genes GATA-3, FOXA1, CK19, integrin-α 6, Elf5 and epithelial cell adhesion molecule (EpCAM) could be performed (Varley et al., 2014). CK19 expression can be detected in almost all samples with small exceptions in cell lines, triple negative, uninvolved ER^+ and triple negative samples. In different cell lines and ER^+ probes the CK19 gene expression is high and in the last three sample groups low (figure 4.47A). The gene expression of integrin-α 6 shows a different pattern: the expression is weaker compared to CK19 and it is mainly expressed in some cell lines, uninvolved ER^+ tissue as well as in healthy and uninvolved triple negative tissue (figure 4.47B). Another interesting marker is Elf5, which is expressed in mammary luminal progenitor cells. This protein is involved in alveolar progression (Oakes et al., 2008)(Oakes et al., 2008). In the analysed data set Elf5 gene expression is even weaker than integrin-α 6 expression. It is almost not expressed in ER^+ samples but in TN, uninvolved ER^+, healthy and uninvolved TN tissues. Few cell lines do express Elf5 but on a very weak level (figure 4.47C). EpCAM is involved in cell signalling, differentiation and migration processes and due to its potential to up-regulate c-myc it is categorized as oncogenic. It is expressed in several carcinomas, therefore functions as a tumour marker, and in progenitor cells. EpCAM is highly

expressed in almost all cell lines and TN samples, weakly expressed in the other tested samples figure 4.47D). GATA-3 expression is very high in the ER^+ tumour samples but weak in TN, healthy and uninvolved TN samples. Few cell lines express GATA-3 as well some uninvolved ER^+ samples (figure 4.47E). In contrast to GATA-3 is FOXA1 expression in general weaker. The FOXA1 expression is mainly in breast cell lines and ER^+ samples observed. A decreased gene expression is found in uninvolved ER^+ and TN and healthy samples. The TN samples have no or a very low expression of FOXA1 (figure 4.47F).



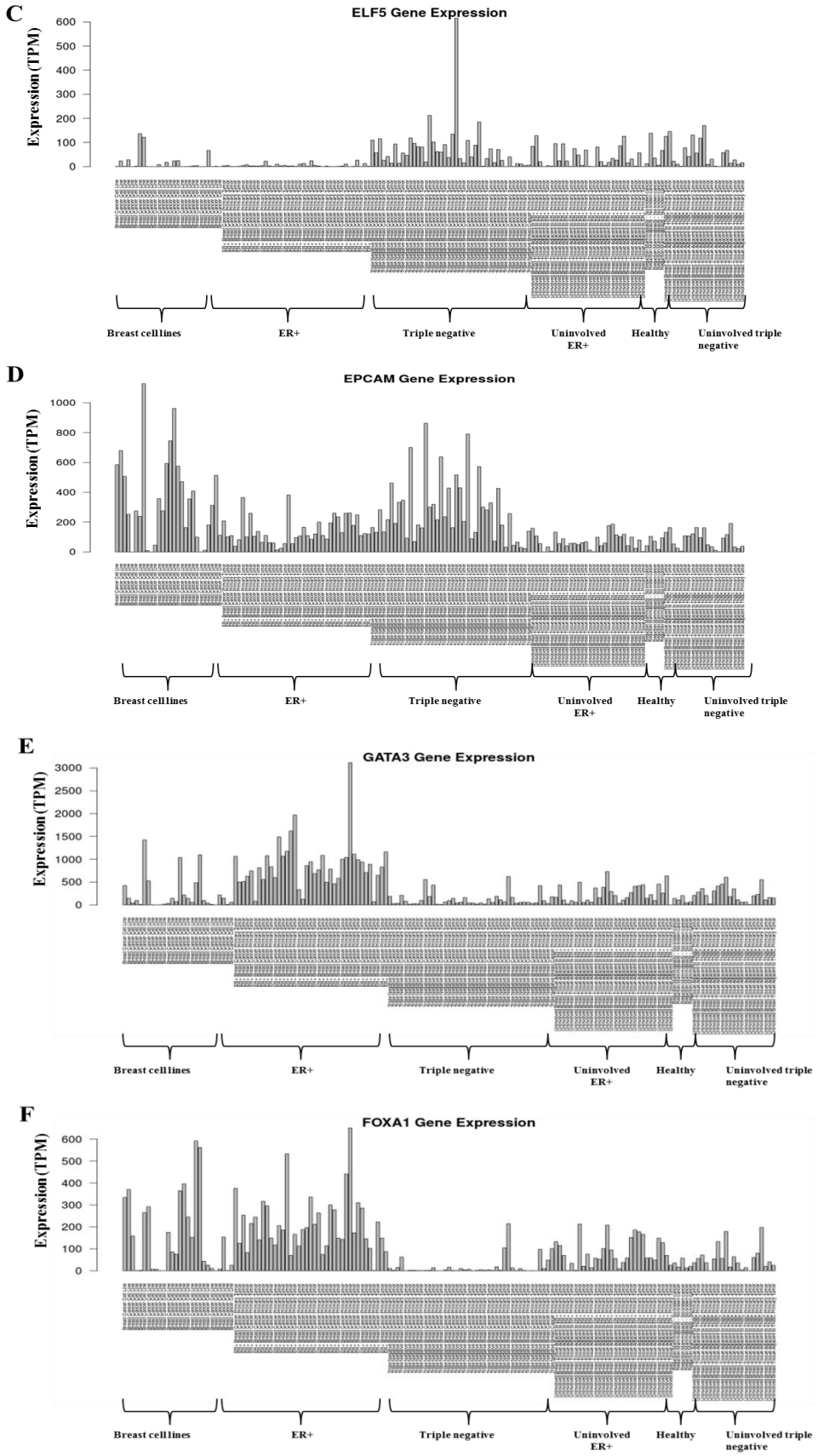


Figure 4.47: Gene expression analysis of A: CK19, B: integrin-α 6, C: Elf5, D: EpCAM, E: GATA-3, F: FOXA1 in breast cell lines, healthy, ER^+ and TN samples as well as the adjacent tissues of ER^+ and TN samples using the RNA-seq data of Varley et al.

4.6.4 *In silico* analysis of NY-BR-1 methylation status

As described before methylation can have an impact on transcription initiation and therefore on gene expression. By using the TCGA database the methylation pattern of the twelve single fragments in the NY-BR-1 promoter region and gene body were analysed in the different molecular subtypes of breast cancer. The location of the fragments is shown in figure 4.48. One fragment (cg20211778) is located within the gene body; all other fragments are located before the transcription start site.

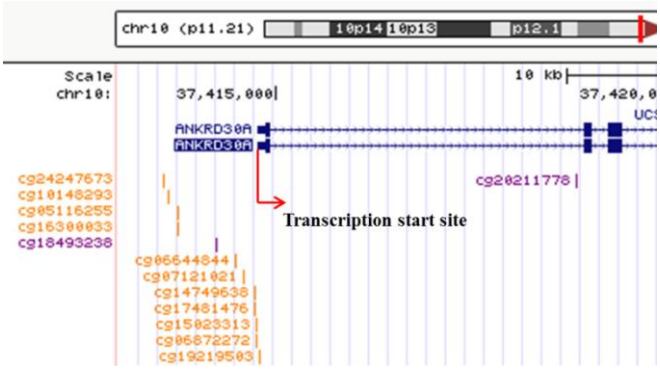


Figure 4.48: Overview of the location of the fragments used in the methylation analysis. Orange: methylated (score >= 600), Purple: partially methylated (200 < score < 600)

It is noticeable that fragment cg06872272, fragment cg06644844 and fragment cg15023313 are highly methylated nearly throughout all samples. The fragment cg20211778 is highly demethylated in HER2, luminal A and B samples. In normal samples this fragment seems to be hyper-methylated (figure 4.49).

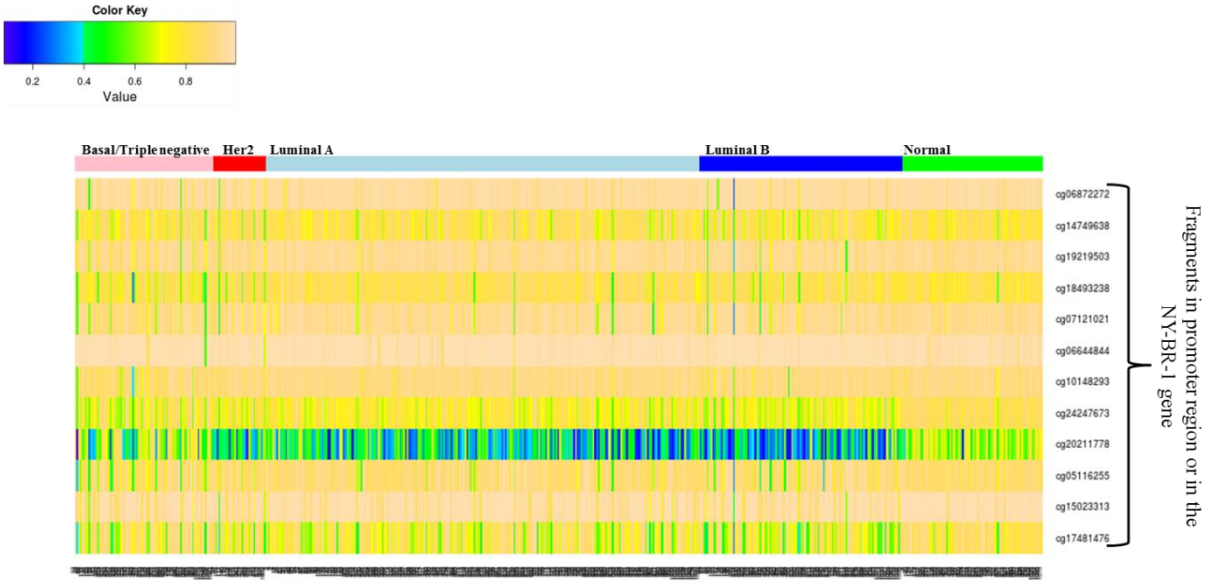


Figure 4.49: Clusterheatmap of the methylation analysis of single fragments located in the promoter region and gene body of NY-BR-1 in the different molecular subtypes of breast cancer. (Colourkey: light yellow: highly methylated, dark blue: highly demethylated)

Due to *in silico* analyses it was possible to identify 69 damaging and 39 splicing SNPs. Moreover, it was shown that NY-BR-1 is mainly expressed in ER^+ breast cancer and less in TN. NY-BR-1 belongs to the top 50 genes beside PGR and ER, which are down-regulated in TNBC. Another analysis with Chip-Seq data revealed that several ER binding sites are present within the NY-BR-1 gene. Gene expression analysis with RNA–Seq data showed that specific genes are expressed in ER^+ breast cancer as well. Additionally, no uniform methylation pattern of NY-BR-1 in the different molecular subtypes could be observed.

5 Discussion

The breast cancer associated antigen NY-BR-1 was discovered by a SEREX screening. NY-BR-1 has restricted expression pattern in breast, prostate and testis, it is over-expressed in the majority (>70 %) of breast cancer patients, humoral and cellular immune response were detected and it has a subcellular localization to the cytoplasm and plasma-membrane. Therefore, it represents a promising target protein for cancer immunotherapy, e.g. antibody-based or CAR therapies or for vaccination because CD4 and CD8 T-cell epitopes were identified (Gardyan et al., 2015; Jäger et al., 2005).

However up to date, the biological function of NY-BR-1 in normal and tumorigenic breast tissues, its interaction partners and how it is transcriptionally regulated are unknown. Therefore the aims of this project were to reveal the transcriptional regulation, the effect of NY-BR-1 over-expression on cellular processes (e.g. proliferation, apoptosis) and protein interaction partner by combining wet lab techniques with *in silico* analyses.

5.1 Transcriptional regulation of NY-BR-1

The transcriptional regulation of gene expression is a complex process including chromatin remodelling, transcription factors, methylation in the promoter region or in the gene itself, and specific promoter elements (e.g. TATA box, enhancer, insulator, silencer, and activators).

To understand how NY-BR-1 is transcriptionally regulated in normal breast tissue and breast tumours, a bioinformatic analysis of the predicted promoter region was performed. With the Alibaba 2.0 tool transcription factor binding sites as well as CpG islands (figure 4.6) were predicted in the assumed promoter region of the ANKRD30A gene (Grabe, 2002). The prediction included binding sites for the nuclear receptors ER, GR, PR, RXR/RAR and T3R but no binding sites were predicted for androgen receptor (AR). AR belongs to the nuclear receptor family and can be activated by binding of the ligand testosterone (Roy et al., 1999). The AR is related to the PR and can be blocked by high dosages of progestins (Bardin et al., 1983). This shows that the actions of the single nuclear receptor are intertwined with each other.

The breast tissue is a dynamic tissue and hormonal regulated underlying the hormonal changes during the menstrual cycle, pregnancy or menopause. Therefore it is supposable that NY-BR-1 expression is regulated by hormones and their receptors acting as TFs. As published by Theurillat et al. (2007), NY-BR-1 expression was compared with ER, HER2 status, EGFR, prognosis and HLA class I antigen expression. 1.444 primary breast cancer

samples, 88 recurrence samples, 525 lymph nodes and 91 distant metastases were analysed by tissue microarrays. The results of the study showed that NY-BR-1 and ER expression can be correlated to each other ($P < 0.0001$) but not NY-BR-1 to HER2 status or EGFR. The co-expression of NY-BR-1 and HLA class I antigen in 15 % of the primary breast tumours and in 6 % of the distant metastases emphasizes once more that NY-BR-1 is a suitable target for cancer immunotherapy. The outcome for breast cancer patients was better when NY-BR-1 was expressed (Theurillat et al., 2007). It is noteworthy to mention that in this study no discrimination between the single molecular subtypes of breast cancer was done but the results indicate that ER influences NY-BR-1 expression.

As numerous binding sites for nuclear hormone receptors were identified, established cell lines were treated with different hormones and analysed for NY-BR-1 expression. For these experiments, NY-BR-1 negative as well as weakly NY-BR-1 positive normal breast and breast cancer cell lines were selected to investigate whether the respective treatment leads to the induction of NY-BR-1 transcription. As shown exemplary for MCF-7 and ZR-75-1 cells, these previous stimulation studies did not lead to an up-regulation of NY-BR-1 (figure 4.7). The measured copy numbers were low and no significant changes in NY-BR-1 expression levels were detectable. Many breast cell lines were analysed regarding their endogenous NY-BR-1 expression but with the available tools no NY-BR-1 expression could be detected on protein level by Western blot analysis. The immortalized cell lines have the advantage to overcome the state of senescence and divide continuously with the right culture conditions compared to the primary cells. But due to the high turn over rate of cell divison during culturing the cells acquire mutations which can impair important cell signalling pathways. However, the *ex vivo* primary material enables to examine the natural state under controlled conditions for a certain time period. The microenvironment in primary tissues is preserved as well as in pleural effusions because the latter were cultured with conditioned medium, containing the cleared supernatant of the effusion including cytokines, chemokines, hormones and growth factors.

Because no suitable cell culture system was available the transcriptional regulation of endogenous NY-BR-1 was investigated in primary cells, normal breast tissues, isolated epithelial cells derived from normal breast tissues and tumour cells from pleural effusions.

5.1.1 Transcriptional regulation of NY-BR-1 in normal breast tissue

To study the transcriptional regulation in normal breast tissue, tissue pieces derived from breast reductions of healthy patients were used and stimulated with different hormones and agents. Before the stimulation experiments could start, the tissues needed to be dissected to purge the glandular tissue from fat and connective tissue. A challenge was to differentiate between connective tissue (in breast a high amount is present) and glandular tissue by eye. This led to a high heterogeneity between the single tissue pieces and thus also between the amount of NY-BR-1 expressing cells. As shown in figure 4.4 the NY-BR-1 expression in normal breast tissue is mosaic-like. Using IHC staining of the stimulated cryo-preserved tissues the heterogeneity of the tissues could have been resolved and the effects of stimulation would have been visible on protein level. But due to technical problems (difficulties to get the tissue section on a slide due to fat and connective tissue, tissue section was disrupted after staining) it was not possible to generate stained tissue sections of the stimulated samples.

Two controls were chosen: tissue pieces without any stimulation cultured as long as the stimulated samples and never cultured tissue pieces. The stimulation within the tissue pieces showed that, in most of the cases, the non-cultured cells have a higher NY-BR-1 expression on mRNA level compared to the tissues/cells being in culture but have never been stimulated (figure 4.8). It seems that the NY-BR-1 expression is lost during the culturing process due to the absence of specific substances.

The tissue pieces were stimulated with specific agents alone or in combination. The function of the used agents is summarized in table 3.4. The results of the stimulated tissue cells showed that none of the tested substances has the ability to have consistently the same effect on NY-BR-1 expression (figure 4.9A). An exception is the stimulation with VitD3 because it suppresses NY-BR-1 expression in all tissue pieces. VitD3 binds to the vitamin D receptor (VDR), which belongs to the superfamily of nuclear receptors regulating the synthesis of many genes, by acting as a ligand-activated transcription factor (Evans, 1988). During puberty, pregnancy and lactation the VDR expression is the highest and is exclusively expressed in the differentiated luminal epithelial cells. It does not co-localize with Ki-67 positive cells suggesting that VDR positive cells are differentiated cells with no proliferative functions (Narvaez et al., 2014; Santagata et al., 2014; Zinser et al., 2002). Moreover, there is a connection between ER and VDR: ER^+ cells have higher VDR expression levels compared to ER^- cells and after application of oestrogen VDR is up-regulated in breast cancer cell lines (Buras et al., 1994; Byrne et al., 2000). As shown by IHC staining NY-BR-1^+ cells can be

ER^+ as well as ER^- in normal tissues (figure 4.39). It needs to be analysed how the VDR expression level conduct in the different stimulated samples because VitD3 stimulation blocks NY-BR-1 expression in the stimulated tissue pieces whereas E2 stimulation has opposed effects. Due to the complexity of VitD3 processing it is difficult to study the VDR effects in the *in vivo* situation of human mammary epithelial cells. In combination with 5'Aza VitD3 stimulation shows conflicting results on NY-BR-1 expression. This might be due to the heterogeneity of the tissue and the different amount of ER^+ cells and this can influence the NY-BR-1 expression. The analysis of the VDR levels in the samples can give information about how diverse its expression pattern is and whether this needs to be considered.

Two patients were also stimulated with other agents (progesterone, retinoic acid, TSA) (figure 4.10). In both analysed patients (HD-T-203/HD-T-258) retinoic acid stimulation decreases NY-BR-1 expression compared to the control whereas the combination with 5'Aza has opposed effects. The retinoic acid receptor (RAR) and retinoic X receptor belong to the nuclear receptor superfamily of ligand-dependent transcription factors (Chambon, 1996). After binding of retinoic acid (9-cis) to RAR a conformational change is induced affecting other proteins, leading to repression or induction of transcriptional activation (Evans and Mangelsdorf, 2014). Further analyses are necessary to reveal the molecular consequences of retinoic acid stimulation. The analysis of the methylation status shows whether different methylation settings in the promoter region of NY-BR-1 are influences the effects of retinoic acid actions on NY-BR-1 expression.

The progesterone stimulated tissue pieces of HD-T-203 and HD-T-258 show that NY-BR-1 expression is down-regulated whereas the combination Aza/Progesterone leads to an up-regulation of NY-BR-1 (figure 4.10). Progesterone is a steroid hormone and is involved in many biological processes such as mammary gland development and reproduction (Macias and Hinck, 2012). The isoform of the progesterone receptor (PR) PR-B drives mammary gland development and expansion (Conneely et al., 2001). The actions and involvements of progesterone and its receptor are manifold: Via paracrine signalling progesterone can induce proliferation to steroid receptor negative cells but it needs to be considered that the PR is expressed just on 10-15 % of the luminal cells (Brisken et al., 1998). Furthermore, it is known that progesterone is a key driver for mammary gland stem cell expansion and maintenance (Brisken, 2013) through the PR-RANKL pathway (Tanos et al., 2013). The PR can bind to progesterone response elements located in the DNA to initiate and alter transcription by recruiting cofactors (Cicatiello et al., 2004; Diep et al., 2015). Moreover, the actions of PR are depending on ER and it is able to "sense" the cell context of presence or absence of activated

pathways (Diep et al., 2015). Thus the interplay between ER and PR is an important aspect to understand a part of the hormonal regulation in the female breast. It seems that for NY-BR-1 expression and regulation both hormones and their receptors are necessary but how this might work in detail needs to be analysed. To test the effect of ER and PR on NY-BR-1 expression an experimental set up can be the stimulation of tissue pieces with E2 and progesterone, E2 and RU-486 (an progesterone receptor antagonist), Tamoxifen and progesterone and control tissues (not stimulated). The NY-BR-1 expression can be analysed on mRNA level via qPCR and on protein level either via IHC staining or Western blot analysis.
In general it was observed that older patients (~ 45+ years) have more NY-BR-1 expression compared to the younger ones (18-30 years). An explanation is that the breast tissue is a dynamic tissue and the genetic profile of specific cell populations are changing due to different hormone levels appearing within a woman's life time such as menstrual cycle, none, one or more pregnancies, breast feeding and menopause (Medina, 2005). These processes force cells to change their normal behaviour because new structures (alveogenesis) need to emerge but also to regress when hormonal changes occur.

To receive a more homogenous cell population the next step was to isolate and culture epithelial cells derived from breast reductions. This cell population consists mainly of myoepithelial and luminal cells. The same controls as with the tissue pieces were taken and analysed regarding their NY-BR-1 expression. The same effect, that NY-BR-1 expression is higher in the non-cultured cells than in the cultured ones, was observed in both analysed patients (figure 4.11). With the isolated epithelial cell populations different effects of certain stimulation agents could be detected compared to the cultured tissue pieces. Patient HD-T-292 and 311 show an increase of NY-BR-1 expression after progesterone stimulation. The same was observed after retinoic acid and TSA were applied. In one patient (HD-T-292) a clear down-regulation of NY-BR-1 was observed with Tamoxifen and Aza/Tamoxifen compared to the control. The second patient (HD-T-311) showed a down-regulation with the combination Aza/Tamoxifen but not with Tamoxifen alone (figure 4.12). Tamoxifen is an ER antagonist and while binding to the ligand-binding domain of ERα a conformational change is induced. Thereby, corepressors are recruited and oestrogen is not able to bind to its receptor anymore leading to a non-proliferative effect (Hayes and Lewis-Wambi, 2015). The conflicting results of Tamoxifen on NY-BR-1 expression could be explained with the observation that depending on the cell type and the present promoter settings 4-hydroxytamoxifen, which functions as antagonists at EREs, can behave as agonists through

indirect pathways (ligand-dependent activation, recruitment of co-factors). Moreover, it has been shown that ERβ can oppose the actions of Tamoxifen on an AP-1 reporter gene or on the SP1 pathway (Kushner et al., 2000; Paech et al., 1997; Saville et al., 2000). AP-1 and SP1 have both predicted binding sites in the promoter region of NY-BR-1 (table 4.1). The stimulation with oestrogen alone or in combination with 5'Aza has conflicting effects. It can be observed that the ER related stimulations (E2, Tamoxifen alone or in combination with 5'Aza) have opposed effects on the NY-BR-1 expression in the two analysed patients. To confirm whether oestrogen and Tamoxifen have a general impact on NY-BR-1 expression in healthy isolated epithelial cells more patients need to be stimulated and analysed.

5.1.2 Transcriptional regulation of NY-BR-1 in pleural effusion cells

Due to early detection methods (mammography) and new therapy strategies (neoadjuvant therapy) breast carcinomas are smaller in size and thus less tumor tissue is available. Therefore, pleural effusions represent a good source to study NY-BR-1 expression in primary tumour cells because the effusions contain tumour cells, immune cells like macrophages and lymphocytes as well as cytokines, chemokines, and growth factors. The diagnosis of the primary tumour is shown in table 4.3. It must be pointed out that the molecular subtype forms another aspect that is important regarding NY-BR-1 expression. Considering the *in silico* analysis NY-BR-1 is mainly expressed in ER^+ breast carcinomas and less in triple negative (figure 4.43). Pleural effusion cells were tested regarding their HER2, EpCAM, CK19 and intra- and extra-cellular NY-BR-1 expression by FACS analysis (figure 4.17). Additionally the mRNA levels of ER, PR and HER2 expression were tested (figure 4.18). The results show that the receptor status of the obtained pleural effusion cells can have changed compared to the one of the primary tumour. This indicates that due to therapy and selection pressure the gene expression profile is changing and this can have consequences on NY-BR-1 expression with either repressing or enhancing it. It needs to be considered that the patients of the obtained pleural effusions differ in therapy status and treatment.

The two controls were also analysed separately (figure 4.13). As opposed to normal breast tissue and isolated normal epithelial cells, the NY-BR-1 expression in NY-BR-1^+ pleural effusions could be maintained for several weeks when cultured with the specific conditioned supernatant (figure 4.19). The NY-BR-1 expression decreases from day zero to day eleven but after this status quo is achieved the NY-BR-1 expression is detectable on mRNA level at least until day 28. The decrease of NY-BR-1^+ cells can be due to that these cells do not proliferate and are going into a senescence state and are therefore overgrown of NY-BR-1- cells.

These observations lead to the conclusion that the NY-BR-1 expression might depend on certain substances existing in the medium. A systematic cytokine and chemokine analysis would help to compare the supernatants of NY-BR-1 negative vs. NY-BR-1 postive, e.g. a multiplex assay containing 50 different cytokines/chemokines. Additionally, the supernatants of NY-BR-1^{+} and NY-BR-1^{-} pleural effusions can be analysed by using 2D gelelectrophoresis or mass spectrometry.

The pleural effusion cells of four patients were additionally stimulated with VPA, Tamoxifen and both in combination with 5'Aza (figure 4.15). Tamoxifen and Aza/Tamoxifen up-regulate NY-BR-1 in two patients (HD-A-73, -78) but in HD-A-88 a down-regulation is seen compared to the control cells. This shows again that NY-BR-1 regulation/expression is intertwined somehow with ER. In three patients (HD-A-73, -78, -88) a down-regulation of NY-BR-1 could be observed by VPA stimulation and in two patients (HD-A-73, -88) in lesser content by Aza/VPA as well. The down-regulation by VPA was unexpected because the HDAC inhibitor VPA is responsible to open the chromatin structure to make a gene more accessible. One of the latest stimulation experiments with TSA and VPA was a time kinetic to see whether the effects of VPA or TSA are changing over time (figure 4.16). The pleural effusion cells were obtained from the same patients known as HD-A-88. The results show that E2, progesterone, VitD3, retinoic acid and VPA up-regulate NY-BR-1 expression compared to the controls. This finding is a contrast to the stimulation of the HD-A-88 pleural effusion cells because the stimulation with the indicated hormones promoted a NY-BR-1 down-regulation. An explanation is that the cells of HD-A-88 were cultured and stimulated over a longer time period (up to six days), whereas the cells of HD-A-90 were cultured for one day and stimulated for twelve hours. TSA alone has not the ability to up-regulate NY-BR-1 in such content as the other substances. At the 8 and 10 h time point TSA/E2 and VPA/TSA/E2 even down-regulate NY-BR-1 expression. Also the other combinations TSA with progesterone, VitD3, retinoic acid and VPA show less NY-BR-1 up-regulation compared to the control and VPA in combination with the mentioned substances. Thus it seems that proteins such as TFs are blocked or repressed by TSA or VPA stimulation that affects NY-BR-1 expression. TSA and VPA inhibit class I and II deacetylases. It is known that HDACs are not just affecting the acetylation status of histones but also non-histone effector proteins such as transcription factors and signal transduction mediators to alter, enhance or repress transcription initiation (Drummond et al., 2005). The following non-histone proteins are influenced by HDAC inhibitors because they increase the acetylation status: GATA-1 (Boyes

et al., 1998), ERα (Wang et al., 2001), androgen receptor (AR) (Fu et al., 2003), YY1 (Yao et al., 2001) and MyoD (Sartorelli et al., 1999). Almost all of these proteins have predicted binding sites in the promoter region of NY-BR-1 (table 4.1). Additionally, tumour associated proteins are influenced by HDAC inhibitors. HER2/neu is down-regulated (Scott et al., 2002) but c-myc (Sasakawa et al., 2003) and ERα (Yang et al., 2000) are up-regulated after applying HDAC inhibitors. In contrast to Yang *et al*., who also took the methylation status of the DNA into account, the group of Reid found out that ERα expression is blocked in MCF-7 cells after incubation with TSA and VPA due to a reduction of mRNA encoding for ERα. Additionally, they detected a reversible promoter shutoff of ERα, pS2 and cyclin D1 (Reid et al., 2005). It might depend on the context that VPA and TSA have the ability to up- or down regulate NY-BR-1 expression. It needs to be testified whether the influence on acetylation on the before described proteins can be mediated by VPA and TSA as well. But if these named proteins are regulated by acetylation it might be possible that they are able to affect NY-BR-1 expression. This needs to be clarified by a systemic analysis of the TFs, hormone receptor expression and control genes to ensure that the appropriate pathways are functional.

In HD-A-66, -68, and -86 NY-BR-1 expression is extremely low. The explanation for this finding can be that these cells had no NY-BR-1 expression at all due to a molecular subtype, which is not known for NY-BR-1 expression, or the therapy application represses NY-BR-1 expression. The receptor status of the pleural effusion cells of these three patients is as followed: HD-A-66/-68 are triple negative and HD-A-86 is weakly $HER2^{+}$ but ER^{-}/PR^{-}. These results confirm the *in silico* analysis of the Varley et al. data set, where NY-BR-1 is expressed in ER^{+} breast cancer but not in triple negative (Varley et al., 2014). With the applied stimulation setting it was not possible to force NY-BR-1 expression in any measurable way. If it turns out to be true that NY-BR-1 is expressed on specific progenitor subsets in the mammary gland than the analysed pleural effusion cells of the patients HD-A-66, -68, and -86 had lost the characteristic features of certain cancer progenitor cells. Maybe when NY-BR-1 negative cells get back into stem/progenitor cell character they would start to express NY-BR-1 in a more frequent manner.

For the future settings several points need to be addressed:

1. Stimulation of tissue pieces and isolated epithelial cells from the same patient to analyse the effect on NY-BR-1 expression.
2. Culturing of tissue pieces and isolated epithelial cells with conditioned medium to maintain NY-BR-1 expression over a longer culture period.

3. Longer stimulation time period (applying the stimulation agents every day for up to three to six days).
4. Other combinations like e.g. E2/progesterone, TSA/E2/progesterone, E2/RU-486, Tamoxifen/progesterone and VPA/E2/progesterone can be tested to detect the influence on NY-BR-1 expression in normal and tumourigenic tissues.
5. A detailed phenotypic characterization of the cells (receptor, TFs) on mRNA and protein level by qPCR and Western blot.
6. Another trial could include the concomitant inhibition of HDACs and DNA methyltransferases (DNMT) to see whether NY-BR-1 can be up-or down-regulated as it is observed with the RARβ2 (Mongan and Gudas, 2005).
7. Using Chip-Seq technology to fish specific TFs and analysing their binding activity in the NY-BR-1 promoter.

5.1.3 Epigentic regulatory mechanisms

Transcriptional regulation not only occurs on histone level but also on the DNA self by de- or hypermethylation in specific promoter regions. From some of the stimulated probes (isolated epithelial and pleural effusion cells) DNA was isolated to analyse the influence of stimulation agents on the methylation status and whether the methylation status can be correlated to NY-BR-1 expression (figure 4.21). Unfortunately, the NY-BR-1 promoter region and the gene body right after transcription start site contain a lot of repetitive elements, which makes it difficult to obtain clear results and to study all interesting regions. This is the reason why it was only possible to analyse the promoter region up to approximately 1.300 bp of transcription start site and to have just two amplicons (number 57, 58) located within the gene.

Considering these results it shows that the differences in methylation status might occur randomly and might not influence NY-BR-1 expression at first sight. It is interesting that in some amplicon areas an increased methylation pattern can be detected but that does not seem to influence the NY-BR-1 expression at this point. To verify these findings more stimulated samples need to be analysed. It is worth to mention that 5-Aza-2´Deoxycytidine, a demethylation agent, functions as a positive control and should result in hypo-methylation of proliferating cells. As the primary cultured cells do not proliferate or proliferate very slowly, Aza treatment may not have had the desired effect and has thus no impact on NY-BR-1 expression. Moreover, it would be interesting to investigate the methylation within the gene to see whether a specific pattern can be observed between stimulated tumour cells and healthy

breast cells. A detailed database analysis of the methylation status of various fragments in the promoter region and gene body of NY-BR-1 is described in chapter V *in silico* analysis. A detailed database analysis of TCGA samples showed that also no uniform methylation pattern of the promoter region of NY-BR-1 can be connected to NY-BR-1 expression in the different molecular subtypes of breast cancer.

Nevertheless, from these results the conclusion can be drawn that NY-BR-1 expression is not mainly regulated by the methylation status in the promoter region. The transcriptional regulation of NY-BR-1 can be influenced by ER and PR but that needs to be confirmed with further experiments. The effects of VPA and TSA stimulation suggest that transcription initiation of NY-BR-1 occurs on histone level and through factors regulated or affected by acetylation.
Considering the limitation of patient samples and obtained material it is difficult to analyse systematically a complex process such as the transcriptional regulation of a protein. It becomes important that many factors (microenvironment of normal and tumourigenic tissue, molcecular subtypes, interplay between TFs, epigenetic modifications) influence the transcriptional initiation of a protein. If further material is acquired the experiments as mentioned before will be performed.

5.2 Functional analysis of NY-BR-1

One major task during this project was to determine the function of NY-BR-1 in normal breast tissue as well as in breast cancer. Knowing the biological function of the tumour-associated antigen NY-BR-1 is important to understand its role in cellular processes. As NY-BR-1 is over-expressed in the majority of breast cancers, the effect of NY-BR-1 over-expression on proliferation, apoptosis and the cell cycle was investigated in transiently transfected normal HEK293 and MCF10A cells, the cancer cell line MCF-7 and transformed HEK293T cells. A challenge was that the transfection efficacy varies strongly between the cell lines and in transiently transfected cells NY-BR-1 expression decreases rapidly after three to four days after transfection (HEK293: 11.6 – 0.397 %, MCF-7: 16.7 – 1-89 %) (figure 4.24). By using FACS analysis it was still possible to analyse low amounts of NY-BR-1^{+} cell populations. The compensation of the single used colours was difficult to adjust because the expressed GFP in the pc3 GFP transfected cells (control cells) was too strong and beamed into the other channels. Therefore, the amount of the pc3 GFP plasmid was reduced for transfection.

The proliferation assay revealed that NY-BR-1$^+$ cells do not proliferate as indicated by the PKH26 labelling and do not go into apoptosis on an increased level (Annexin/7-AAD staining) compared to the non-transfected cells, pc3 GFP transfected cells and NY-BR-1$^-$ cell population (figure 4.22-4.28).

The effect of NY-BR-1 over-expression in transfected cell lines was analysed in more detail by performing cell cycle analyses. The cell cycle analysis of pc3 NY-BR-1 transfected cells from day one to day three showed that NY-BR-1$^+$ cells arrest more in the G1 phase than the non-transfected cells and the NY-BR-1$^-$ population cells (figure 4.28-4.31). These results indicate that NY-BR-1 is responsible to slow down the cell cycle and prevents the cells going from G1 to S phase.

It is known that the retinoblastoma (rb), p107 and p130 repressors bind to the E2F transcription factor and thereby preventing the transition. Several cell cycle checkpoints are present. The G1 phase checkpoint is necessary to sense any type of DNA damage. If the decision is made to go into the S phase the level of cyclin D rises and forms complex with diverse proteins. Translocation of the E2F4 and E2F5 repressors is taking place as well as transcriptional activation of downstream targets to promote the G1 to S transition (Skotheim et al., 2008). In the case that NY-BR-1 promotes G1 phase arrest a detailed analysis of the different phosphorylated Rb proteins and cyclin D1 is necessary. Ki-67, a proliferation marker, could also be an indicator. If NY-BR-1 positive cells are in G1 arrest they should be Ki-67 negative because this protein is absent in resting cells. IHC staining of normal mammary gland tissue showed that NY-BR-1$^+$ cells are Ki-67$^-$ indicating that the NY-BR-1$^+$ cells are in a quiescent state.

One major problem was that the cells quickly loose the NY-BR-1 expression. A reason for this phenomenon can be that the cells digest the plasmid and remove the already expressed NY-BR-1 because it slows down the proliferation. It was tried to generate stable transfected cell lines but the cells proliferate very slowly and after two months of cultivation the cells lost NY-BR-1 expression on RNA and protein level. A way to circumvent this problem is to generate viral transduced cells stably expressing NY-BR-1 and repeat the proliferation assay and cell cycle analysis. Another explanation of the G1 phase arrest of NY-BR-1$^+$ cells is the assumption that NY-BR-1 is expressed in stem/progenitor cells of the mammary gland. This kind of cells is in a quiescent state and they start proliferation after a certain stimulus. This will be illustrated in detail in the stem/progenitor chapter (chapter 5.4).

To analyse the differences between NY-BR-1 positive vs. NY-BR-1 negative cells laser micro dissection (LMD) is a useful tool. That tool enables to cut out single cells or tissue areas

positive or negative for NY-BR-1. The drawback of this method is that enough material needs to be obtained which can be difficult due to low amount of tissues. Moreover for RNA-Seq analysis the RNA quality is poor because the laser cutting is performed at room temperature. Nevertheless, with the obtained material it is possible to do proteomics or DNA analysis.

5.3 Identification of NY-BR-1 protein interaction partner

Protein-protein interactions are regulated by a particular stimulus or signal and depend on cell-type, developmental stage of the cell, cell cycle phase and the presence of other proteins. The identification of NY-BR-1 protein interaction partner helps to understand the role of NY-BR-1 in cellular processes (proliferation, cell cycle etc.) in a given cellular environment since execution of a particular protein function will be strongly dependent on contact with surfaces of neighboring proteins (Ngounou Wetie et al., 2014).

To find specific interaction partners of NY-BR-1 a Co-IP with a following mass spectrometry analysis was performed with three transiently pc3 NY-BR-1 transfected cell lines (HEK293, MCF-7, MCF-10A). Co-IPs of lysates from primary cells were also done with α-NY-BR-1 (C-18) and α-NY-BR-1#2 to analyse the interaction partners of NY-BR-1 in a natural setting but after silver - or coomassie blue staining no NY-BR-1 at the appropriate height in the SDS-PAGE gels could be detected.

For analysis the unique proteins in the NY-BR-1 transfected cells were taken and compared with each other in all cell lines. Differences between the NY-BR-1 scores/significant protein sequences between the three cell lines could be observed and this can be due to a difference in transfection efficacy resulting in different protein expression levels and thus the scores are varying. Furthermore, the differences of efficiency in immunoprecipitation need to be considered although all samples were treated the same way. As stated by Ngounou Wetie *et al.* this biochemical method for protein-interaction detection has some weaknesses: the used polyclonal antibody (α-NY-BR-1 (C-18)) can lack high specificity and avidity and therefore a risk is of detecting non-specific interactions, due to the presence of highly abundant contaminating proteins, exists. The monoclonal antibody α-NY-BR-1#2 was used to perform Co-IPs of either pc3 NY-BR-1 transfected cell lines or of lysates derived from primary material but after silver staining no NY-BR-1 band could be detected at the appropriate height. However, the analysis of large molecular weight complexes can be a problem. In some cases genetic modifications prior to the biochemical analysis need to be performed as well as the destruction of the native cellular environment (Ngounou Wetie et al., 2014). Although this method has some disadvantages, it can give hints for potential interaction

partner in short time, is cost-effective, ease-of-use, and is compatible with other downstream methods.

Tubulin beta-4B chain (TBB4) is the only protein shared between all cell lines and this protein is a major component of microtubules, which are involved in chromosome separation during mitosis/meiosis and cell trafficking. Moreover, microtubules are a component of the cytoskeleton of every cell and of the mitotic spindles (Vicente and Wordeman, 2015). The protein K220, also known as KRT76, belongs to the type II keratin family and could be detected in MCF-10A and MCF-7 cells. This protein constitutes the intermediate filaments of the cytoskeleton and help to maintain cell function and structure. The third interesting protein is SNX27, which belongs to the sorting nexin family. Proteins of this family are involved in endocytosis of plasma membrane receptors and it supports their protein trafficking. SNX27 protein contains a PX domain (phospholipid-binding motif) and a PDZ domain (protein-protein interaction domain that is often found in the postsynaptic density of neuronal excitatory synapses). This domain combination is unique to SNX27 (Wang et al., 2014; Worby and Dixon, 2002). The HS71L protein could be detected in two lanes of the MCF-10A and MCF-7 cell lines. This heatshock protein is involved to maintain the right folding of a newly synthetized protein (chaperone character) and to prevent aggregation of existing proteins. The identified proteins, which might be potential interaction partner of NY-BR-1, can be clustered as components of the cytoskeleton, and regulation of protein trafficking and folding.

It is known that NY-BR-1 localizes in the cytoplasm and plasma-membrane (Seil et al., 2007). The translation of mRNA into a protein takes place in the cytosol of a cell or across the membrane of the endoplasmatic reticulum. The synthesized peptide is stored in the endoplasmatic reticulum until it gets packed into a vesicle and transported to its final location(Reece et al., 2014). Therefore it is not unlikely that NY-BR-1 interacts with proteins involved in trafficking and folding. Further analyses by western blot or a Co-IP against TBB4 and K220 can help to understand how these two proteins are influenced by NY-BR-1 expression.

5.4 NY-BR-1 expression in stem/progenitor cells in the mammary gland

As described in the stem cell chapter 1.5 many different stem/progenitor cell populations are known to be present in the mammary gland. They fulfil specific functions depending on the developmental state of the breast (puberty, pregnancy, breast feeding, and menopause). Although many questions about stem/progenitor cells in the breast remain unclear it is

assumed that the different molecular subtypes of breast cancer arise due to these different stem/progenitor cell populations. The theory describes that due to mutations occurring in the normal different stem/progenitor cell populations the subtypes emerge (Reya et al., 2001). The evidence for this theory is that breast cancer stem cells (BCSC) share similarities with normal stem cells regarding their self-renewal and differentiation (Ginestier et al., 2007; Ponti et al., 2005). A second theory assumes that BCSCs arise from epithelial-mesenchymal transition (EMT) because the cells, undergone EMT, have an increased ability to form mammospheres and they express many genes associated with that process (Mani et al., 2008; Velasco-Velázquez et al., 2012). As described in chapter 1.5 the different molecular subtypes of breast cancer are prognostic and require unique therapy strategies and thus it is important to characterize the cells as detailed as possible.

The NY-BR-1 expression pattern in the normal breast is very mosaic-like and only single cells per gland express this protein as assessed by IHC staining of various normal breast tissue sections (figure 4.4). The NY-BR-1 expression in breast cancer depends on the molecular subtype and mainly ER^+ specimens express NY-BR-1 (figure 4.49). Moreover, the proliferation assay and the cell cycle analyses revealed that NY-BR-1 over-expressing cells do not proliferate, do not show increased apoptosis and accumulate in the G1 phase. All these results led to the assumption that NY-BR-1 might be a potential stem/progenitor cell marker in the mammary gland.

As proposed by Dontu *et al.*, mammary gland progenitor cells can form non-adherent mammospheres with the ability to differentiate along all three mammary epithelial lineages (myoepithelial cells that form the basal layer of ducts and alveoli, ductal epithelial cells that line the lumen of ducts, and alveolar epithelial cells that synthesize milk proteins). Secondary and tertiary mammospheres are more enriched with a 100 % of multilineage progenitors compared to the primary spheres containing these progenitor cells to a lesser content (Dontu et al., 2003).

Mammospheres from four different healthy patients were generated with isolated epithelial cellsfrom reduction mammoplastiesaccording to the protocol of Dontu *et al*. A lot of cells are needed to generate enough spheres to do all analyses including immunofluorescence and IHC staining, Western blot, qPCR and to generate secondary and tertiary spheres. That is the reason why not all analyses were carried out with all patients. Nevertheless, it was possible to generate primary, secondary and tertiary spheres of one patient and the qPCR results show that NY-BR-1 is expressed in the mammospheres, but expression is lost over the culturing time period suggesting that beside growth factors (basic fibroblast growth factor (FGF), EGF)

other factors (e.g. E2, progesterone) are needed to regulate and maintain NY-BR-1 expression in culture.

Compared to the mammospheres NY-BR-1 expression can be maintained over a period of several weeks in pleural effusion cells if incubated with the conditioned medium but the expression decreases until it is diminished completely (figure 4.19). It is important to characterize the substances (cytokines, chemokines, hormones), which are implied in the supernatant of the pleural effusion because it is very likely that they help to maintain NY-BR-1 expression under the requirement that the cells of the patients are NY-BR-1 positive. If these substances can be identified by e.g. a bioplex analysis and applied to the spheres it might be possible to maintain NY-BR-1 expression even over a longer culture period, which would improve the analysis of the gene expression profile. The NY-BR-1 expression on mRNA level varies between the four patients and this can be due to different kinds of stem/progenitor cell subsets or difference in the amount of progenitor cells (figure 4.35). To get a better impression, which proteins are up- or down-regulated while NY-BR-1 is expressed or not, HER2, ER, PR, GATA-3, FOXA1, and integrin-α 6 expressions were analysed via qPCR. Results of other groups showed that specific progenitor cells in the mammary gland are ER and PR negative and this could be confirmed by the four analysed patients (Lim et al., 2009; Visvader and Stingl, 2014). Strikingly, the HER2 expression was high compared to the other tested receptors/proteins. It belongs to the epidermal growth factor family besides HER1, HER3 and HER4 and is a transmembrane protein. HER2 has no own ligand and becomes activated after HER2/HER3 dimerization (Koutras and Evans, 2008). HER2 is known for enhancing cell proliferation and repressing apoptosis (Olayioye, 2001) and shows thus opposed effects compared to NY-BR-1 expression. But it is noteworthy that the HER2 expression was analysed in primary cells and the NY-BR-1 over-expression effect occurred in transfected cell lines.

As shown by Theurillat *et al.* NY-BR-1 expression could not be linked to HER2 status in breast cancer samples (Theurillat et al., 2007). Now it is necessary to test whether the HER2 expression on mRNA level is translated to protein level in the mammospheres and whether they correlate with NY-BR-1 expression. It needs to be considered that this HER2 expression can be a coincidence because HER2 and HER3 are needed for mammary gland development at different stages e.g. for alveogenesis (Stern, 2003). Due to missing information it is not known whether the tested patients were pregnant and had breast feeding.

However, NY-BR-1 expression is not associated with other progenitor cell marker, such as GATA-3, FOXA1, and integrin-α 6. Dontu *et al.* did a transcriptional profiling of

mammosphere derived cells and their analysis revealed that GATA-3 expression is only occurring in differentiated cells (Dontu et al., 2003). It is suggested that the transcription factors GATA-3 and FOXA1 interact with each other and that they are also known to be regulators of ERα expression. GATA-3 is involved in the formation of terminal end buds at puberty and luminal cell differentiation (Kouros-Mehr et al., 2006). It is a driver of breast oncogenesis and low GATA-3 expression is described for ER/PR negative, HER2 over-expressing breast cancer (Mehra et al., 2005). FOXA1 binds to specific promoters of genes associated with regulating signalling pathways and the cell cycle (Carlsson and Mahlapuu, 2002). By binding to cis-regulatory regions of heterochromatin it enhances the binding between ERα and chromatin (Laganière et al., 2005). The interplay between GATA-3, FOXA1 and ER are thought to be needed for normal mammary gland development, particularly the differentiation of mammary stem/ progenitor cells to the ER-positive lineage (Hisamatsu et al., 2014). Due to the absence of ER in all tested NY-BR-1 expressing mammospheres and a weak GATA-3/FOXA1 expression it is likely that NY-BR-1 is expressed in an undifferentiated stem/progenitor cell subset in the mammary gland.

Integrin-α 6 (CD49f) is used for the characterization of the myoepithelial lineage in the mammary gland (Shackleton et al., 2006; Stingl et al., 2006). In the tested mammospheres integrin-α 6 is expressed but it cannot be connected to a specific NY-BR-1 expression profile suggesting it is not the best link to NY-BR-1 expression. This assumption is confirmed by the gene expression profile of integrin-α 6 of the RNA-Seq data set published by Varley et al. It shows that integrin-α 6 is rather expressed in certain breast cell lines, healthy samples and in ER^+ and TN adjacent breast tissue but not in ER^+ and TN breast cancer tissues while NY-BR-1 is mainly expressed in ER^+ breast cancer samples (figure 4.43) (Varley et al., 2014).

The immunofluorescent staining of mammospheres of two different patients shows that the spheres are NY-BR-1^+. The spheres of one patient (HD-T-258) were stained with CK14 (myoepithelial cell marker) and the spheres of the other patient (HD-T-311) stained with CK18 (luminal cell marker). But the expression of both markers is very weak until negative suggesting that progenitor cells expressing NY-BR-1 belong to a much undifferentiated subset of stem/progenitor cells in the healthy mammary gland. In contrast to the mammaospheres the gene expression analysis of the Varley *et al*. data set demonstrates that CK19 expression is increased in ER^+ breast cancer samples compared to TN, healthy, ER^+ and TN uninvolved tissues. These results show that one has to distinguish between mammospheres and tumorigenic tissue and that NY-BR-1 expression occurs in luminal epithelial cells in breast cancer samples.

To characterize the gene expression profile of NY-BR-1 positive mammospheres a microarray could help to understand the up- or down-regulation of certain genes when NY-BR-1 is expressed to get more details which progenitor cell population expresses NY-BR-1. Another option can be the culturing of potential progenitor cells derived from mammospheres on collagen or matrigel and analysing if mammary gland structures are develop after adding of e.g. prolactin (Dontu et al., 2003). If so they could be tested on NY-BR-1 expression and maybe the function of NY-BR-1 becomes clearer.

Co-localization of NY-BR-1 with ER in serial tissue sections of normal mammary gland tissues, analysed by IHC staining (figure 4.39), show that NY-BR-1$^+$ cells can be either ER$^+$ as well as ER$^-$. From these data it can be concluded that NY-BR-1 is expressed in mammospheres but it seems that it must be an early undifferentiated subset with ER$^+$/ER$^-$ negative status. Furthermore, a detailed gene expression profile is necessary of NY-BR-1 positive vs NY-BR-1 negative cells using e.g. the LMD method. In the case that enough cells are available a FACS sort with additionally RNA-Seq analysis can be performed. Another option is the analysis of co-localization of NY-BR-1 with PR, CD49f, EpCAM, and integrin-α 6 via IHC staining in healthy and tumorigenic breast tissues. The results can give information about the expression profile and can provide evidence of the specific progenitor subset.

5.5 *In silico* analysis of NY-BR-1

Computational analysis becomes increasingly important to understand complex settings such as interactions between proteins, to understand and analyse large data sets (sequencing data) and to find differences in the gene profile of cells in normal but also in tumorigenic tissue. Here, several different analyses were performed regarding the *in silico* characterization of NY-BR-1.

5.5.1 Molecular subtypes

As described before breast cancer is classified according to the different receptor status of ER, PR and HER2. This classifications needs to be done to evaluate the best therapy strategy for the patient. However, the highest NY-BR-1 expression could be detected in ER$^+$ breast cancer (figure 4.44) but here it was not possible to discriminate the ER$^+$ breast cancer samples further into luminal A or B. This would have been helpful to see whether the ER$^+$ breast cancer samples are also HER2$^+$, thus a kind of luminal B because then it could be comparable with the gene expression profile of the mammospheres. These additional data could offer valuable clues to analyse whether NY-BR-1 expression in normal breast samples is correlated to HER2

expression but not to ER/PR. And these findings could be compared to tumorigenic breast where it seems to be the other way around. As described in the mammosphere chapter GATA-3, FOXA1 and ER are tight together to regulate transcription of specific genes. The gene expression profile of GATA-3 and FOXA1 of the Varley *et al*. data set shows that these two factors are mainly expressed in ER^+ breast cancer samples (Varley et al., 2014). Compared to TN breast cancer NY-BR-1 was significantly over-expressed in ER^+ samples indicating that the receptor status is important and a marker for NY-BR-1 expression (figure 4.44). Moreover, NY-BR-1 belongs to the top 50 genes down-regulated in TN breast cancer (table 4.11). It is even more down-regulated than ER encouraging that NY-BR-1 has important functions in the mammary gland and is regulated by at least ER.

The methylation pattern of the stimulated samples, described in chapter 5.1, could not be correlated to a specific NY-BR-1 expression pattern. But here it needs to be considered that just a small number of samples could be analysed. Therefore, the information of the TCGA database was consulted because the data sets contain larger sample numbers. As described earlier the NY-BR-1 expression differs in the different molecular subtypes of breast cancer. The breast cancer samples in the TCGA database were grouped according to the known subtypes but a specific methylation pattern of the analysed fragments to NY-BR-1 expression in a specific subtype could not be linked. This result suggests that another regulation mechanism is responsible for NY-BR-1 expression, e.g. on histone level or by ER and PR. To look into a regulation mechanism on histone level Chip-Seq can be used with antibodies against common histone modifications (acetylation, methylation) to see whether differences exist between NY-BR-1 positive and negative cells.

5.5.2 SNP analysis

One of the main analyses included a detailed look at SNPs in the assumed promoter region and within the gene of NY-BR-1. SNPs are the most common type of genetic variation and they contribute to every human disease, conferring susceptibility or resistance or influencing interaction with environmental factors (Li et al., 2014). Genome-wide association studies (GWAS) focus more on SNPs in coding sequences but the non-coding SNPs, found augmented in promoters, enhancers and on 3' ends, can affect the expression of neighbouring genes and can be responsible for specific disease phenotypes (Molineris et al., 2013). 56 SNPs are located in the NY-BR-1 promoter up to 2000 bp upstream of the transcription start site and few are located in predicted binding sites of different TFs (table 4.10). Of great interest are the SNPs located in the GR, HNF-3 (also known as FOXA1/2/3), TBP, SP1,

C/EBPα, YY1, c-Fos, c-Jun and AP-1 predicted binding sites. The functions of the encoded proteins are shown in table 5.1. These TFs are known to be proto-oncogenes or are involved in different cellular processes. If NY-BR-1 is also involved in regulating cellular processes it needs to be evaluated whether certain SNPs have an influence on the binding affinity of specific TFs and if yes which consequences this might have on the transcription initiation of NY-BR-1 and for the cell self. Another important issue is to analyse if specific SNPs are present in breast cancer patients and can be correlated to NY-BR-1 expression and thus also with disease progression. From different studies it is known that SNPs can have impact on patient prognosis, outcome and course of the disease (Chirilă et al., 2014; Yamamoto-Ibusuki et al., 2014).

Table 5.1: Specific TFs affected by SNPs with predicted binding sites in the promoter region of NY-BR-1. Information of the TFs are from www.genecards.org

Transcription factor	**Function**
AP-1 (activator protein 1)	• A heterodimeric protein composed of proteins belonging to the c-Fos, c-Jun, ATF and JDP families. • Regulation of gene expression in response to a variety of stimuli (cytokines, growth factors, stress, and bacterial and viral infections). • It controls a number of cellular processes (differentiation, proliferation, and apoptosis). • AP-1 upregulates transcription of genes containing the TPA DNA response element.
C/EBPα (CCAAT/enhancer-binding proteins)	• Critical transcription factor in adipogenesis, growth control and differentiation of various tissues. • The protein encoded is a bZIP transcription factor binding as a homodimer to certain promoters and enhancers. • It forms heterodimers with the related proteins CEBP-beta and CEBP-gamma. • The protein can interact with CDK2 and CDK4, thereby inhibiting these kinases and causing growth arrest in cultured cells.
c-Fos	• Is a proto-oncogene. • Protein forms heterodimer with c-jun → resulting in the formation of AP-1 complex which binds DNA at AP-1 specific sites at the promoter and enhancer regions of target genes and converts extracellular signals into changes of gene expression (cell proliferation, differentiation and survival; genes associated with hypoxia; and angiogenesis). • It induces a loss of cell polarity and epithelial-mesenchymal transition, leading to invasive and metastatic growth in mammary epithelial cells.
c-Jun	• In combination with c-Fos, forms the AP-1 early response transcription factor. • It is activated through double phosphorylation by the JNK pathway but has also a phosphorylation-independent function. • Overexpression of c-jun in MCF-7 cells → increased aggressiveness, as shown by increased cellular motility, increased expression of a matrix degrading enzyme MMP-9, increased *in vitro* chemo-invasion and tumour formation in nude mice in the absence of exogenous estrogens. • The MCF-7 cells with c-jun overexpression became unresponsive to estrogen and tamoxifen → c-jun overexpression is proposed to lead to an estrogen-independent phenotype in breast cancer cells.
GR (Glucocorticoid receptor)	• The receptor to which cortisol and other glucocorticoids bind, regulates genes controlling the development, metabolism, and immune response.
HNF-3 (also known as FOXA1/2/3)	• FOXA1 (HNF-3α) and FOXA2 (HNF-3β) belong to the forkhead/winged helix DNA binding domain family. • Function of FOXA1 is further described in the text (see 5.4).

SP1 (Specificity protein 1)	• The encoded protein is involved in cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling. • Post-translational modifications such as phosphorylation, acetylation, glycosylation, and proteolytic processing significantly affect the activity of this protein, which can be an activator or a repressor.
TBP (TATA box binding protein)	• General transcription factor that binds specifically to a DNA sequence called the TATA box. • It helps position RNA polymerase II over the transcription start site of the gene.
YY1 (ying-yang 1)	• YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. • It may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter. • YY1 has been shown to interact with Histone deacetylase 2, FKBP3, ATF6, Myc, SAP30, EP300, HDAC3, NOTCH1 and RYBP.

Most of the SNPs found in the gene body of NY-BR-1 are located in introns (figure 4.41). Introns are non-coding regions and are removed during the splicing process but this process can be influenced by intronic SNPs. 39 of the downloaded SNPs are annotated as splice region variants in dbSNP (figure 4.42). If SNPs are located at a donor or acceptor site they can affect the splicing machinery and new transcripts can be produced translated into a protein or it leads to a break off and no protein can be translated (Lewandowska, 2013).

The publicly available data contains only the somatic mutations. The TCGA data base offers the possibility, after an application is requested, to analyse the raw data, such as exome sequences to look at germline mutations. It is of note that it would be interesting to analyse whether certain damaging/splicing SNPs in the NY-BR-1 gene can be correlated with clinical data such as presence in specific molecular subtypes of breast cancer, age, and ethnicity or if they can be related to different transcripts of NY-BR-1.

5.5.3 Splice Variants

An analysis of the RNA-Seq data set from Varley et al showed that in ER^+, TN uninvolved and ER^+ uninvolved tissue the latter exons (exon 20, 26-36) of NY-BR-1 are highly expressed compared to the first ones. The highest coverage can be found in the ER^+ breast cancer samples. In TN breast cancer no difference can be observed between the coverage of all exons-they are all low (figure 4.45) (Varley et al., 2014). These findings suggest that different transcripts exist. In further analysis it needs to be validated if these transcripts are translated into protein or if they have another function or will be degraded with the consequence that no NY-BR-1 is expressed. To identify splice variants a DNA microarray could help. The exon coverage confirms the results of Theurillat *et al*. They performed an *in situ* hybridization of exon 4-7 and 30-33 and observed that the latter are significantly stronger expressed and more frequently detected on mRNA level (70 % vs. 35 %, $p < 0.0001$). They draw the conclusion

that alternative splice variants that lack 5-prime sequences of NY-BR-1 are present, which also could be observed in the normal breast epithelium (Theurillat et al., 2008). For the qPCR analysis a primer pair was used with binding sites in the latter exons (see figure 6.1). It was tried to design appropriate primer for the 5' end but many repetitive elements are existing in the 5' area and due to the existing NY-BR-1.1 isoform it was difficult to get specific primer. To test whether splice variants are existing and whether they have different functions within cellular processes deletion constructs could be used. The cell lines (MCF-7, MCF-10A, HEK293, and HEK293T) could be transiently transfected and analysed by FACS measurements to analyse their effect on proliferation and cell cycle.

5.5.4 Gene analysis of different transcription factors

By refining the characterization of the breast cancer samples a better profile for NY-BR-1 expression is possible. That ER might be crucial for NY-BR-1 expression is also shown by predicted binding sites in the promoter region (table 4.1) and a Chip-Seq analysis where several ER binding sites are present in the NY-BR-1 gene (figure 4.46). These binding sites are all found up to intron 21 and none later suggesting that ER might be involved in the expression of NY-BR-1 and also which exons are expressed. ER regulates biological processes by affecting gene expression. After oestrogen-ligand binding to ER they are able to bind directly to oestrogen response elements (EREs) in promoters of target genes to attract co-regulators to the receptor to change chromatin structure and to promote RNA polymerase II binding. Another way to regulate genes by ER is that ER interacts with other TF complexes like Fos/Jun (AP-1 responsive elements) or SP1 in promoters without EREs. Thus, in general the ER action depends on the composition of co-regulatory proteins present in a cell, which binding elements are existing in a promoter and the hormones because they can modulate transcription and signalling pathways (Heldring et al., 2007; Katzenellenbogen et al., 1996; Nilsson et al., 2001). Considering all these data ER seems to be involved in the regulation of NY-BR-1 expression but how this finally works needs to be sorted out with further experiments.

5.5.5 Progenitor marker

The limitation of RNA-Seq data is that it does not give evidence whether the mRNA is also translated into a protein. Nevertheless, it can be used to get an impression what is transcribed into mRNA.

Several progenitor marker in the mammary gland were chosen. Elf5, belongs to the luminal lineage in the adult mammary gland, is mainly expressed in TN breast cancer and to a lower extent in healthy, ER^+ and TN uninvolved tissues suggesting that Elf5 is down-regulated in luminal originated ER^+ breast cancer. Elf5 also does not play a role to be co-expressed with NY-BR-1 in certain breast cancer types but it needs to be tested whether it can be linked to NY-BR-1 expression in normal mammary gland samples. Beside Elf5 integrin-α 6 is a progenitor marker for the myoepithelial lineage. The expression pattern of integrin-α 6 is comparable with Elf5 because just a weak expression can be found in ER^+ breast cancer samples (figure 4.37) (Varley et al., 2014). The expression pattern of Elf5 and integrin-α 6 underline that NY-BR-1+ cells are originate from luminal epithelial cells in breast cancer.
In silico analysis supports the NY-BR-1 data/results generated by wet lab techniques. Now it is necessary to look into new data sets to confirm the previous findings and to characterize the regulation and co-expressing genes of NY-BR-1 in more detail.

To summarize the results, NY-BR-1 is expressed mainly in well differentiated hormone sensitive estrogen receptor positive breast cancer subtypes and it is likely that NY-BR-1 expression is influenced by ERα and/or PR expression, but the association of NY- BR-1 expression and ER signaling still needs to be elucidated. Furthermore, transiently NY-BR-1 expressing cells show an inhibited proliferation rate and accumulate in the G1 phase. *In vivo*, endogenous NY-BR-1 is expressed in non proliferating (Ki-67 negative) cells in normal breast tissue. The protein potentially interacts with the tubulin beta-4B chain suggesting a crucial role during mitosis. Moreover, NY-BR-1 was shown to be expressed in progenitor cells of the mammary gland. The phenotypic characterization of these progenitor cells and the question whether the protein is also expressed in cancer stem cells is part of ongoing studies.

6 Appendix

Section of NY-BR-1 sequence

Figure 6.1: A section of the NY-BR-1 sequence showing the binding sites of the primer pairs 60/59 and 100/80. The latter were used for qPCR analysis.

DAB-BOND staining protocol

For general IHC stainings the automated BOND system from Leica was used. In the following a general staining protocol for epitope retrieval 1 is described.

Table 6.1: BOND staining protocol for ER 1. (Cryo preserved tissue start here; ** Variable dilution medium; ***Polymeramplifier)*

Program	Time	Temperature
BOND Dewax Solution	30 sec	72°C
BOND Dewax Solution		72°C
BOND Dewax Solution		RT
3 x 99 % EtOH		RT
2 x BOND Wash solution		RT
BOND wash solution	5 min	RT
2 x BOND ER Solution 1 (Citratpuffer 6.0)		RT
BOND ER Solution 1 (Citratpuffer 6.0)	20 min	100°C
BOND ER Solution 1 (Citratpuffer 6.0)	12 min	RT
3 x BOND wash solution		37°C
BOND wash solution*	3 min	RT
Peroxide Block	10 min	RT
3 x BOND wash solution		RT
Serumblock (10 % goatserum in PBS)	10 min	RT
3 x BOND wash solution		RT
Primary antibody in TBS/10 % FBS**	30 min	RT
3 x BOND wash solution		RT
Post primary (secondary antibody, rabbit α-mouse IgG)	8 min	RT
3 x BOND wash solution	2 min	RT
Polymer (secondary antibody, poly***-HRP α-rabbit IgG)	8 min	RT
2 x BOND wash solution	2 min	RT
Deionized water		RT
Mixed DAB Refine		RT
Mixed DAB Refine	5 min	RT
3 x Deionized water		RT
Haematoxylin	5 min	RT
3 x Deionized water		RT
BOND Wash solution	5 min	RT
Deionized water	~	RT

Vector cards

The pcDNA3.1 NY-BR-1 plasmid contains the full length NY-BR-1 sequence and was used for several applications and is shown exemplary.

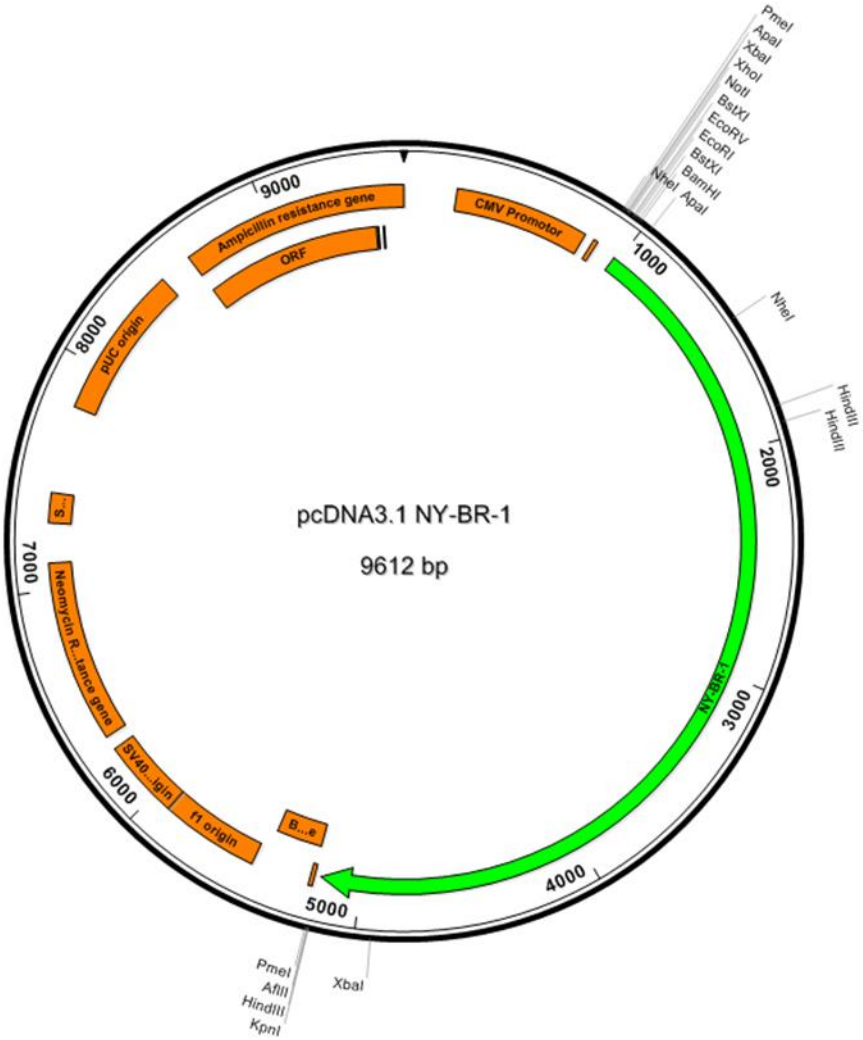


Figure 6.2: pcDNA3.1 NY-BR-1 plasmid used for transfection.

Proliferation Assay

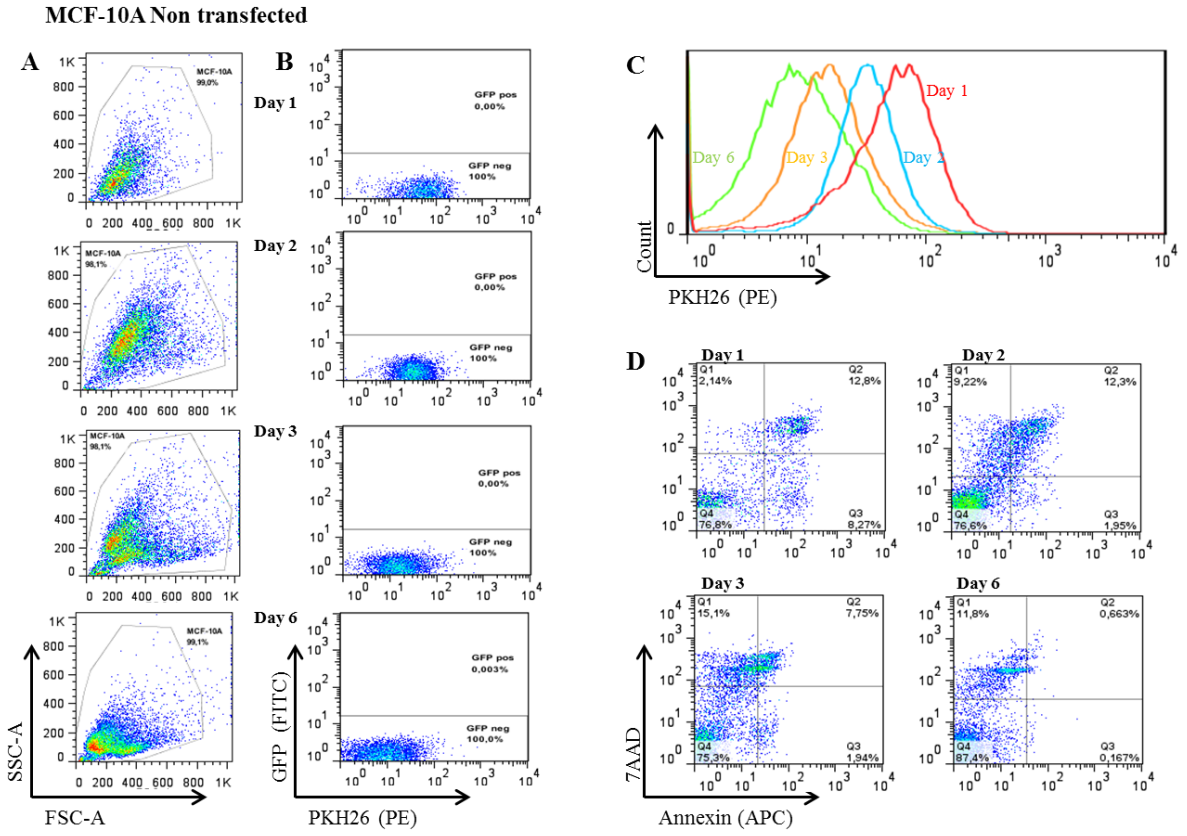


Figure 6.3: Overview of the proliferation assay of MCF-10A non-transfected cells. A: The chosen cell population for further analysis leaving the cell junk out. B: The cell population divided into GFP positive/negative for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

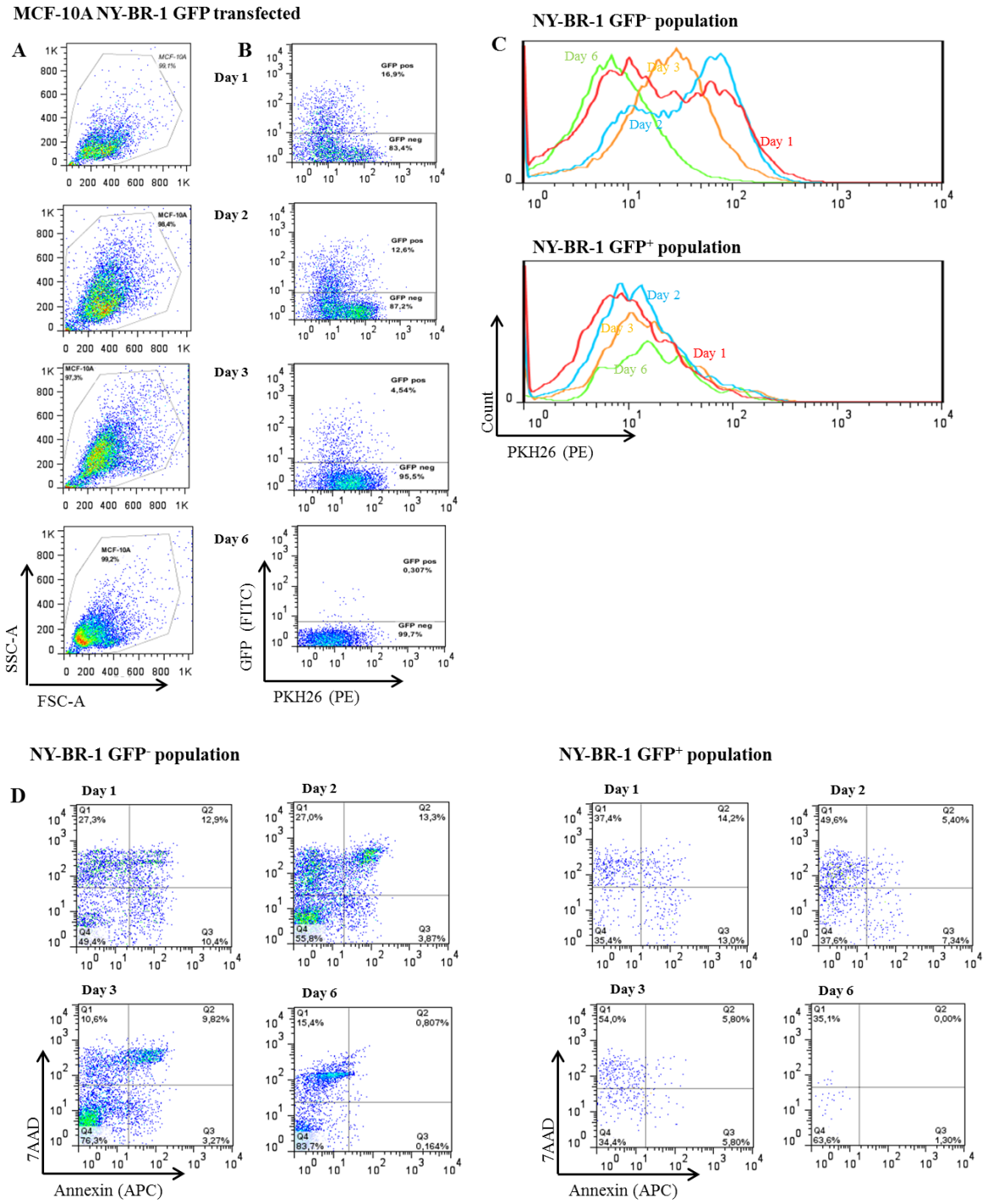


Figure 6.4: Results of the proliferation assay of MCF-10A pc3 NY-BR-1 GFP transfected cells. A: The chosen cell population for further analysis leaving the cell junk out. B: The cell population divided into GFP positive/negative for further analysis. C: Proliferation analysis of the GFP negative cell population versus the GFP positive from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

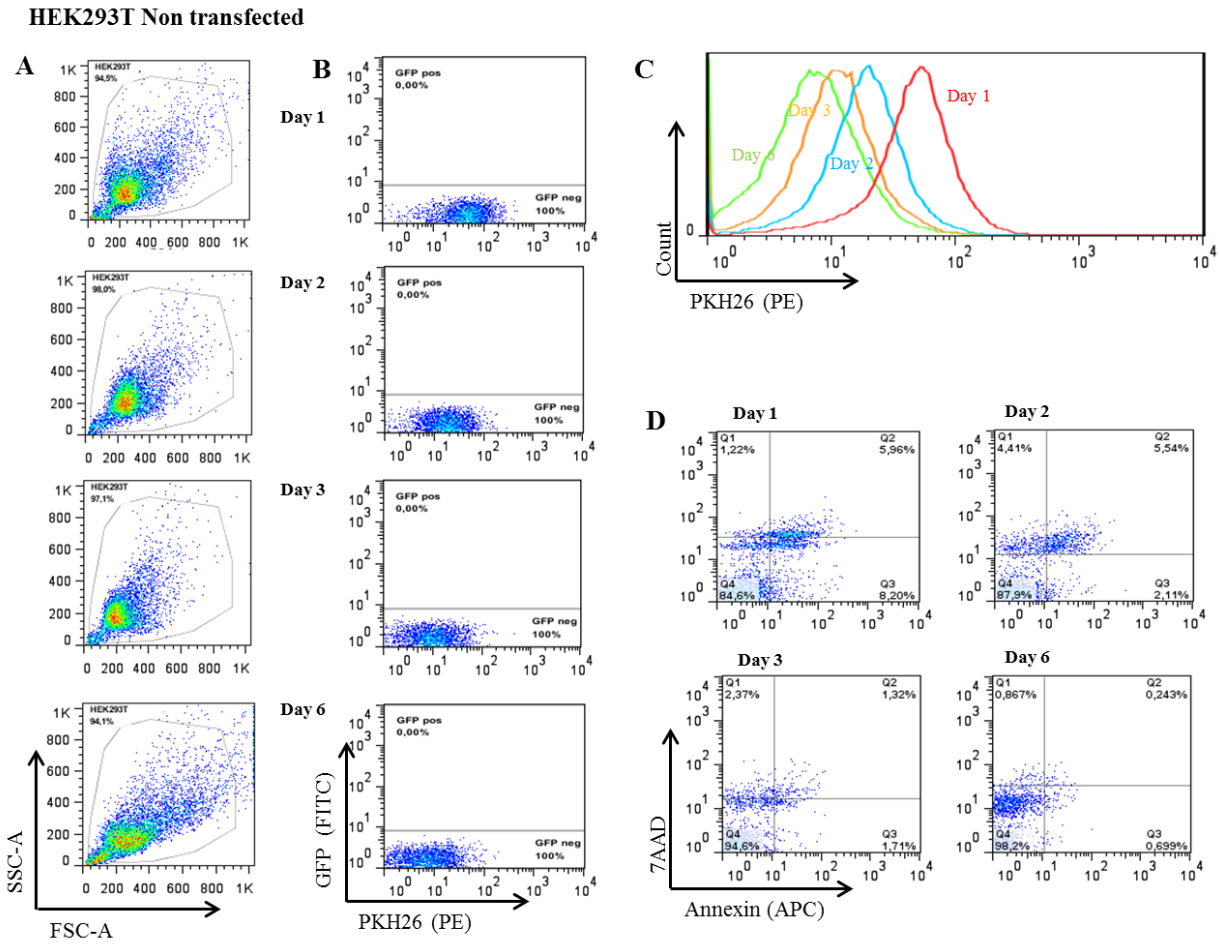


Figure 6.5: Overview of the proliferation assay of HEK293T non-transfected cells. A: The chosen cell population for further analysis leaving the cell junk out. B: The cell population divided into GFP positive/negative for further analysis. C: Proliferation analysis of the GFP negative cell population from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

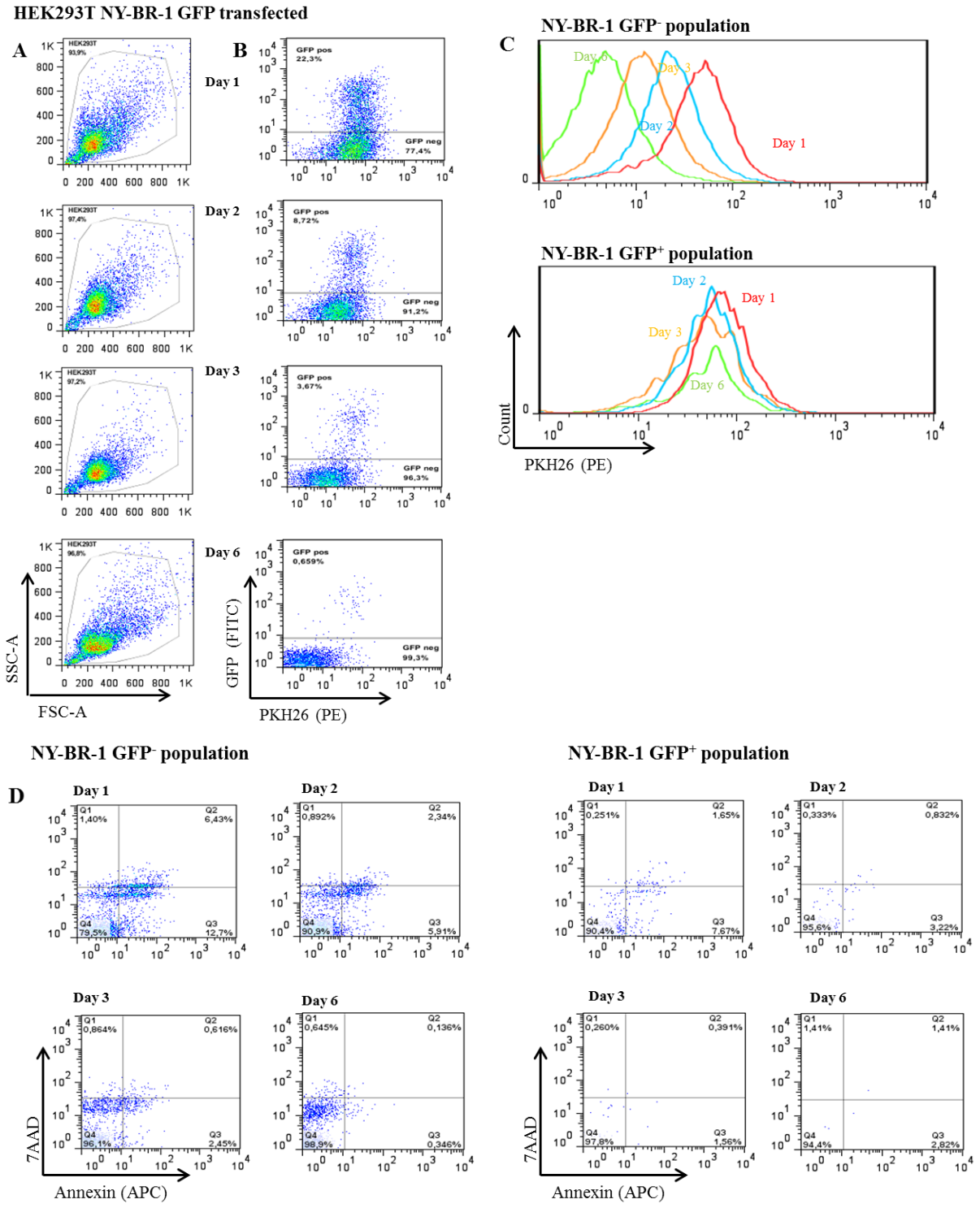


Figure 6.6: Results of the proliferation assay of HEK293T pc3 NY-BR-1 GFP transfected cells. A: The chosen cell population for further analysis leaving the cell junk out. B: The cell population divided into GFP positive/negative for further analysis. C: Proliferation analysis of the GFP negative cell population versus the GFP positive from day one to day six. D: Cell population staining with 7-AAD and Annexin V to discriminate between living/dead and apoptotic cells.

Cell cycle analysis

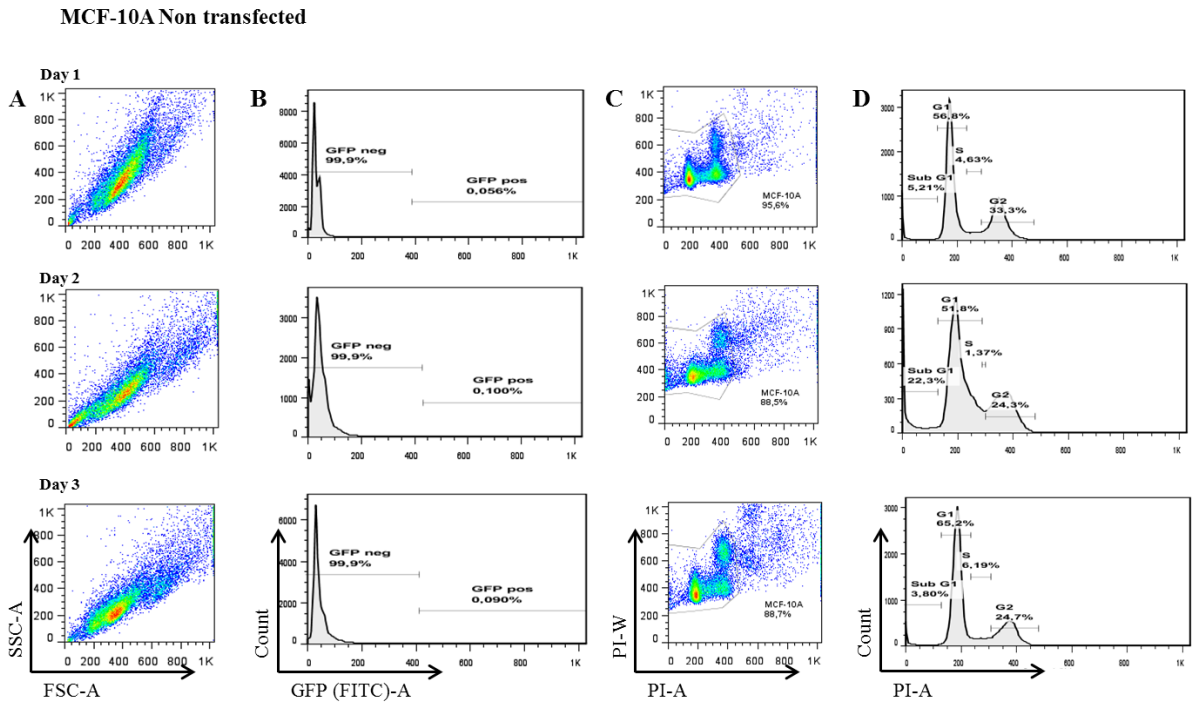


Figure 6.7: Results of the cell cycle analysis of MCF-10A non-transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell junk out. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

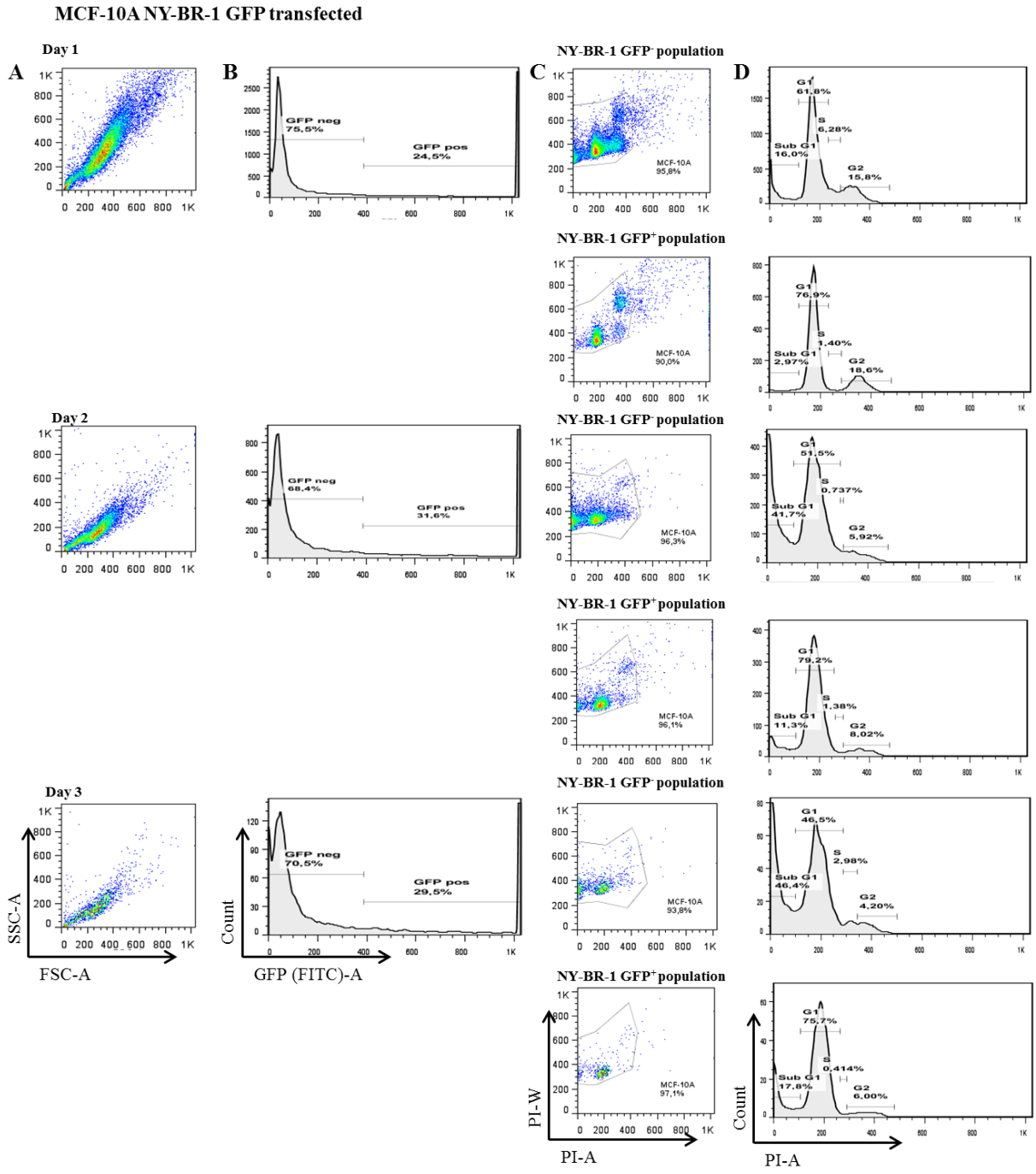


Figure 6.8: Results of the cell cycle analysis of MCF-10A NY-BR-1 GFP transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell junk out in either NY-BR-1 GFP negative or NY-BR-1 GFP positive cell population. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

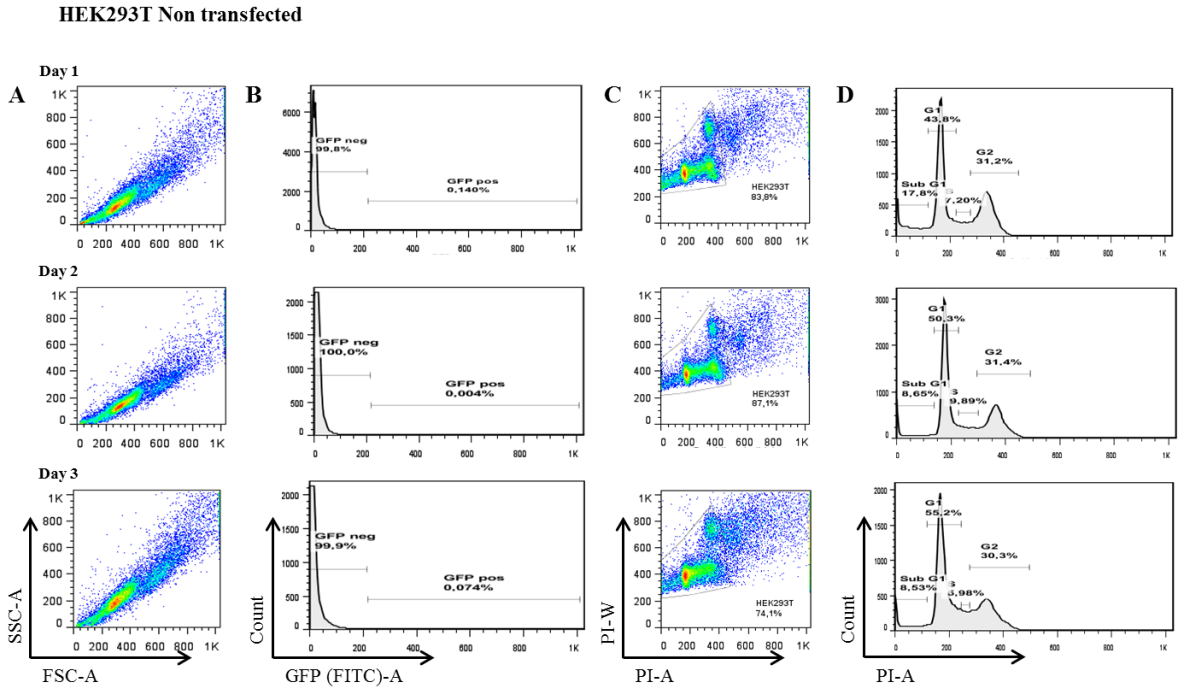


Figure 6.9: Results of the cell cycle analysis of HEK293T non-transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell junk out. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

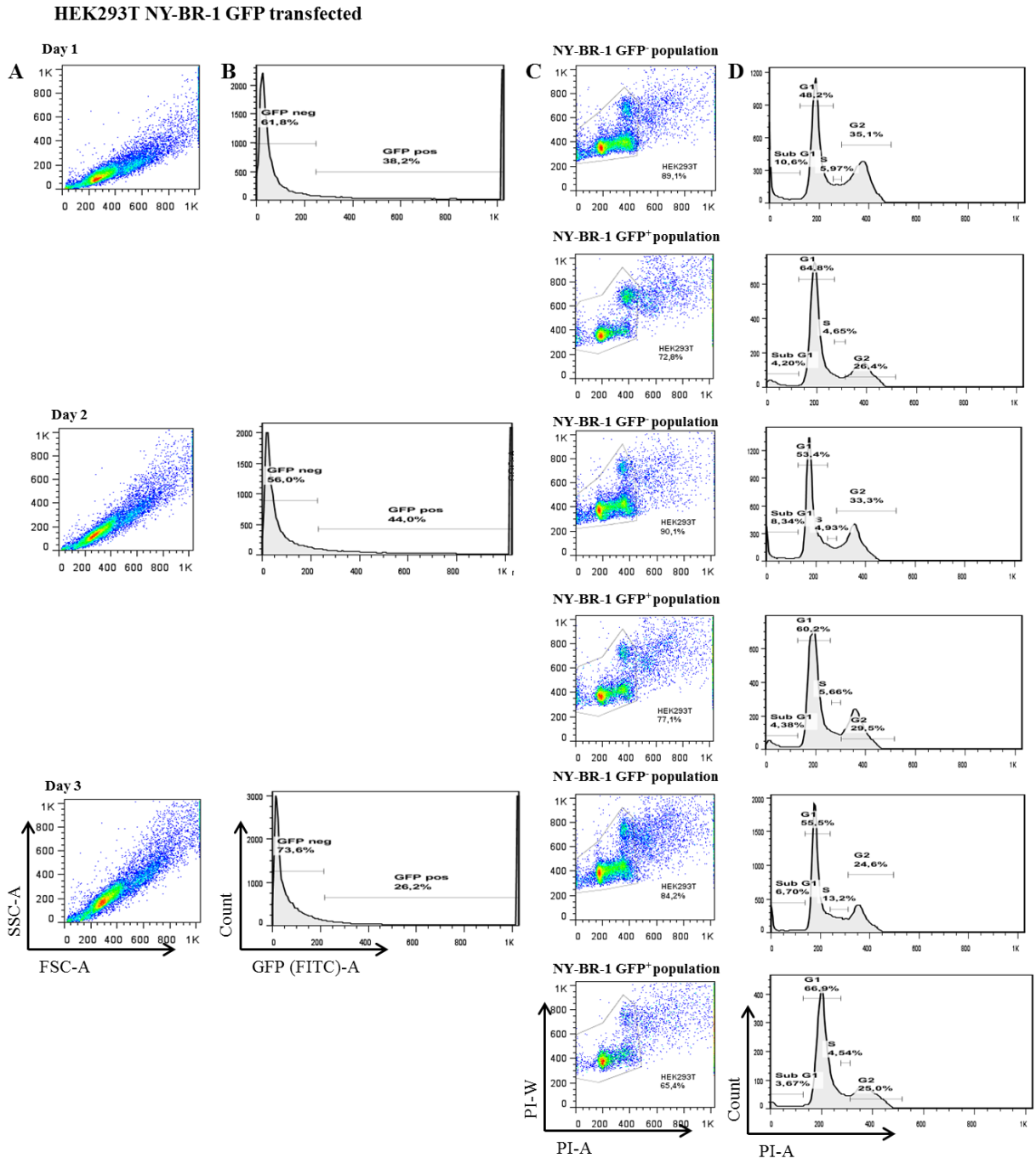


Figure 6.10: Results of the cell cycle analysis of HEK293T NY-BR-1 GFP transfected cells. A: Showing the cell population for further analysis taking all cells into account. B: The cell population divided into GFP positive/negative for further analysis. C: Gates set to leave the duplets and cell junk out in either NY-BR-1 GFP negative or NY-BR-1 GFP positive cell population. D: Cell cycle analysis dividing the cells into subG1, G1, S, and G2 phase after PI staining.

7 References

Agus, D.B., Akita, R.W., Fox, W.D., Lewis, G.D., Higgins, B., Pisacane, P.I., Lofgren, J.A., Tindell, C., Evans, D.P., Maiese, K., et al. (2002). Targeting ligand-activated ErbB2 signaling inhibits breast and prostate tumor growth. Cancer Cell *2*, 127–137.

Ambatipudi, S., Gerstung, M., Pandey, M., Samant, T., Patil, A., Kane, S., Desai, R.S., Schäffer, A.A., Beerenwinkel, N., and Mahimkar, M.B. (2012). Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. Genes, Chromosomes and Cancer *51*, 161–173.

Ambatipudi, S., Bhosale, P.G., Heath, E., Pandey, M., Kumar, G., Kane, S., Patil, A., Maru, G.B., Desai, R.S., Watt, F.M., et al. (2013). Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse. PLoS ONE *8*, e70688.

Amos, S.M., Duong, C.P.M., Westwood, J.A., Ritchie, D.S., Junghans, R.P., Darcy, P.K., and Kershaw, M.H. (2011). Autoimmunity associated with immunotherapy of cancer. Blood *118*, 499–509.

Anderson, D.J., Durieux, J.K., Song, K., Alvarado, R., Jackson, P.K., Hatzivassiliou, G., and Ludlam, M.J.C. (2011). Live-cell microscopy reveals small molecule inhibitor effects on MAPK pathway dynamics. PLoS ONE *6*, e22607.

Aran, D., Toperoff, G., Rosenberg, M., and Hellman, A. (2011). Replication timing-related and gene body-specific methylation of active human genes. Hum. Mol. Genet. *20*, 670–680.

Armstrong, L. (2014). Epigenetics (Garland Science).

Arts, J., de Schepper, S., and Van Emelen, K. (2003). Histone deacetylase inhibitors: from chromatin remodeling to experimental cancer therapeutics. Curr. Med. Chem. *10*, 2343–2350.

Auguste, P., Lemiere, S., Larrieu-Lahargue, F., and Bikfalvi, A. (2005). Molecular mechanisms of tumor vascularization. Crit. Rev. Oncol. Hematol. *54*, 53–61.

Bardin, C.W., Brown, T., Isomaa, V.V., and Jänne, O.A. (1983). Progestins can mimic, inhibit and potentiate the actions of androgens. Pharmacol. Ther. *23*, 443–459.

Bauzon, M., and Hermiston, T. (2014). Armed therapeutic viruses - a disruptive therapy on the horizon of cancer immunotherapy. Front Immunol *5*, 74.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Blanke, C.D., Demetri, G.D., von Mehren, M., Heinrich, M.C., Eisenberg, B., Fletcher, J.A., Corless, C.L., Fletcher, C.D.M., Roberts, P.J., Heinz, D., et al. (2008). Long-term results from a randomized phase II trial of standard- versus higher-dose imatinib mesylate for patients with unresectable or metastatic gastrointestinal stromal tumors expressing KIT. J. Clin. Oncol. *26*, 620–625.

Boisguérin, V., Castle, J.C., Loewer, M., Diekmann, J., Mueller, F., Britten, C.M., Kreiter, S., Türeci, Ö., and Sahin, U. (2014). Translation of genomics-guided RNA-based personalised cancer vaccines: towards the bedside. Br. J. Cancer *111*, 1469–1475.

Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. Nature *396*, 594–598.

Brentjens, R.J., and Curran, K.J. (2012). Novel cellular therapies for leukemia: CAR-modified T cells targeted to the CD19 antigen. Hematology Am Soc Hematol Educ Program *2012*, 143–151.

Brischwein, K., Schlereth, B., Guller, B., Steiger, C., Wolf, A., Lutterbuese, R., Offner, S., Locher, M., Urbig, T., Raum, T., et al. (2006). MT110: a novel bispecific single-chain antibody construct with high efficacy in eradicating established tumors. Mol. Immunol. *43*, 1129–1143.

Brisken, C. (2013). Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. Nat. Rev. Cancer *13*, 385–396.

Brisken, C., Park, S., Vass, T., Lydon, J.P., O'Malley, B.W., and Weinberg, R.A. (1998). A paracrine role for the epithelial progesterone receptor in mammary gland development. Proc. Natl. Acad. Sci. U.S.A. *95*, 5076–5081.

Brümmendorf, T.H., Cortes, J.E., de Souza, C.A., Guilhot, F., Duvillié, L., Pavlov, D., Gogat, K., Countouriotis, A.M., and Gambacorti-Passerini, C. (2015). Bosutinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukaemia: results from the 24-month follow-up of the BELA trial. Br. J. Haematol. *168*, 69–81.

Buras, R.R., Schumaker, L.M., Davoodi, F., Brenner, R.V., Shabahang, M., Nauta, R.J., and Evans, S.R. (1994). Vitamin D receptors in breast cancer cells. Breast Cancer Res. Treat. *31*, 191–202.

Byrne, I.M., Flanagan, L., Tenniswood, M.P., and Welsh, J. (2000). Identification of a hormone-responsive promoter immediately upstream of exon 1c in the human vitamin D receptor gene. Endocrinology *141*, 2829–2836.

Carlsson, P., and Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. Dev. Biol. *250*, 1–23.

Castle, J.C., Kreiter, S., Diekmann, J., Löwer, M., van de Roemer, N., de Graaf, J., Selmi, A., Diken, M., Boegel, S., Paret, C., et al. (2012). Exploiting the mutanome for tumor vaccination. Cancer Res. *72*, 1081–1091.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304.

Chambon, P. (1996). A decade of molecular biology of retinoic acid receptors. FASEB J. *10*, 940–954.

Chang, H.L., Gillett, N., Figari, I., Lopez, A.R., Palladino, M.A., and Derynck, R. (1993). Increased transforming growth factor beta expression inhibits cell proliferation in vitro, yet increases tumorigenicity and tumor growth of Meth A sarcoma cells. Cancer Res. *53*, 4391–4398.

Cheang, M.C.U., Chia, S.K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P.S., Parker, J.S., et al. (2009). Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J. Natl. Cancer Inst. *101*, 736–750.

Chirilă, D.N., Bălăcescu, O., Popp, R., Oprea, A., Constantea, N.A., Vesa, S., and Ciuce, C. (2014). GSTM1, GSTT1 and GSTP1 in patients with multiple breast cancers and breast cancer in association with another type of cancer. Chirurgia (Bucur) *109*, 626–633.

Christiansson, L., Soderlund, S., Mangsbo, S.M., Hjorth-Hansen, H., Höglund, M., Markevarn, B., Richter, J., Stenke, L., Mustjoki, S., Loskog, A., et al. (2015). The Tyrosine Kinase Inhibitors Imatinib and Dasatinib Reduce Myeloid Suppressor Cells and Release Effector Lymphocyte Responses. Mol. Cancer Ther.

Cicatiello, L., Addeo, R., Sasso, A., Altucci, L., Petrizzi, V.B., Borgo, R., Cancemi, M., Caporali, S., Caristi, S., Scafoglio, C., et al. (2004). Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor

(progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. Mol. Cell. Biol. *24*, 7260–7274.

Cobleigh, M.A., Vogel, C.L., Tripathy, D., Robert, N.J., Scholl, S., Fehrenbacher, L., Wolter, J.M., Paton, V., Shak, S., Lieberman, G., et al. (1999). Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. J. Clin. Oncol. *17*, 2639–2648.

Collin, C., Ouhayoun, J.-P., Grund, C., and Franke, W.W. (1992). Suprabasal marker proteins distinguishing keratinizing squamous epithelia: Cytokeratin 2 polypeptides of oral masticatory epithelium and epidermis are different. Differentiation *51*, 137–148.

Conneely, O.M., Mulac-Jericevic, B., Lydon, J.P., and De Mayo, F.J. (2001). Reproductive functions of the progesterone receptor isoforms: lessons from knock-out mice. Molecular and Cellular Endocrinology *179*, 97–103.

Cretu, A., and Brooks, P.C. (2007). Impact of the non-cellular tumor microenvironment on metastasis: potential therapeutic and imaging opportunities. J. Cell. Physiol. *213*, 391–402.

Darzynkiewicz, Z., Halicka, H.D., and Zhao, H. (2010). Analysis of cellular DNA content by flow and laser scanning cytometry. Adv. Exp. Med. Biol. *676*, 137–147.

Diep, C.H., Daniel, A.R., Mauro, L.J., Knutson, T.P., and Lange, C.A. (2015). Progesterone action in breast, uterine, and ovarian cancers. J. Mol. Endocrinol. *54*, R31–R53.

Disis, M.L. (2014). Mechanism of action of immunotherapy. Semin. Oncol. *41 Suppl 5*, S3–S13.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Dorschner, M.O., Amendola, L.M., Turner, E.H., Robertson, P.D., Shirts, B.H., Gallego, C.J., Bennett, R.L., Jones, K.L., Tokita, M.J., Bennett, J.T., et al. (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am. J. Hum. Genet. *93*, 631–640.

Drummond, D.C., Noble, C.O., Kirpotin, D.B., Guo, Z., Scott, G.K., and Benz, C.C. (2005). Clinical development of histone deacetylase inhibitors as anticancer agents. Annu. Rev. Pharmacol. Toxicol. *45*, 495–528.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998.

Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004a). The three Es of cancer immunoediting. Annu. Rev. Immunol. *22*, 329–360.

Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004b). The immunobiology of cancer immunosurveillance and immunoediting. Immunity *21*, 137–148.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. *102*, 15785–15790.

Eirew, P., Kannan, N., Knapp, D.J.H.F., Vaillant, F., Emerman, J.T., Lindeman, G.J., Visvader, J.E., and Eaves, C.J. (2012). Aldehyde dehydrogenase activity is a biomarker of primitive normal human mammary luminal cells. Stem Cells *30*, 344–348.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. Nat. Rev. Genet. *8*, 286–298.

Evans, R.M. (1988). The steroid and thyroid hormone receptor superfamily. Science *240*, 889–895.

Evans, R.M., and Mangelsdorf, D.J. (2014). Nuclear Receptors, RXR, and the Big Bang. Cell *157*, 255–266.

Free, R.B., Hazelwood, L.A., and Sibley, D.R. (2009). Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy. In Current Protocols in Neuroscience, J.N. Crawley, C.R. Gerfen, M.A. Rogawski, D.R. Sibley, P. Skolnick, and S. Wray, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.),.

Fu, M., Rao, M., Wang, C., Sakamaki, T., Wang, J., Di Vizio, D., Zhang, X., Albanese, C., Balk, S., Chang, C., et al. (2003). Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. Mol. Cell. Biol. *23*, 8563–8575.

Fu, N., Lindeman, G.J., and Visvader, J.E. (2014). The Mammary Stem Cell Hierarchy. In Current Topics in Developmental Biology, (Elsevier), pp. 133–160.

Gajewska, M., Zielniok, K., and Motyl, T. (2013). Autophagy in Development and Remodelling of Mammary Gland. In Autophagy - A Double-Edged Sword - Cell Survival or Death?, Y. Bailly, ed. (InTech),.

Gardyan, A., Osen, W., Zörnig, I., Podola, L., Agarwal, M., Aulmann, S., Ruggiero, E., Schmidt, M., Halama, N., Leuchs, B., et al. (2015). Identification of NY-BR-1-specific CD4(+) T cell epitopes using HLA-transgenic mice. Int. J. Cancer *136*, 2588–2597.

Gillison, M.L., Chaturvedi, A.K., and Lowy, D.R. (2008). HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. Cancer *113*, 3036–3046.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567.

Goldhirsch, A., Winer, E.P., Coates, A.S., Gelber, R.D., Piccart-Gebhart, M., Thürlimann, B., Senn, H.-J., and Panel members (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann. Oncol. *24*, 2206–2223.

Grabe, N. (2002). AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol. (Gedrukt) *2*, S1–S15.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U.S.A. *100*, 3983–3988.

Harries, M., and Smith, I. (2002). The development and clinical use of trastuzumab (Herceptin). Endocr. Relat. Cancer *9*, 75–85.

Hayes, E.L., and Lewis-Wambi, J.S. (2015). Mechanisms of endocrine resistance in breast cancer: an overview of the proposed roles of noncoding RNA. Breast Cancer Res. *17*, 542.

Heemskerk, B., Kvistborg, P., and Schumacher, T.N.M. (2012). The cancer antigenome. The EMBO Journal *32*, 194–203.

Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Ström, A., Treuter, E., Warner, M., et al. (2007). Estrogen receptors: how do they signal and what are their targets. Physiol. Rev. *87*, 905–931.

Higano, C.S., Schellhammer, P.F., Small, E.J., Burch, P.A., Nemunaitis, J., Yuh, L., Provost, N., and Frohlich, M.W. (2009). Integrated data from 2 randomized, double-blind, placebo-controlled, phase 3 trials of active cellular immunotherapy with sipuleucel-T in advanced prostate cancer. Cancer *115*, 3670–3679.

Hisamatsu, Y., Tokunaga, E., Yamashita, N., Akiyoshi, S., Okada, S., Nakashima, Y., Taketani, K., Aishima, S., Oda, Y., Morita, M., et al. (2014). Impact of GATA-3 and FOXA1 expression in patients with hormone receptor-positive/HER2-negative breast cancer. Breast Cancer.

Holopainen, T., Bry, M., Alitalo, K., and Saaristo, A. (2011). Perspectives on lymphangiogenesis and angiogenesis in cancer. J Surg Oncol *103*, 484–488.

Huertas, D., Soler, M., Moreto, J., Villanueva, A., Martinez, A., Vidal, A., Charlton, M., Moffat, D., Patel, S., McDermott, J., et al. (2012). Antitumor activity of a small-molecule inhibitor of the histone kinase Haspin. Oncogene *31*, 1408–1418.

Jäger, D., Stockert, E., Güre, A.O., Scanlan, M.J., Karbach, J., Jäger, E., Knuth, A., Old, L.J., and Chen, Y.T. (2001). Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. Cancer Res. *61*, 2055–2061.

Jäger, D., Karbach, J., Pauligk, C., Seil, I., Frei, C., Chen, Y.-T., Old, L.J., Knuth, A., and Jäger, E. (2005). Humoral and cellular immune responses against the breast cancer antigen NY-BR-1: definition of two HLA-A2 restricted peptide epitopes. Cancer Immun. *5*, 11.

Jäger, D., Filonenko, V., Gout, I., Frosina, D., Eastlake-Wade, S., Castelli, S., Varga, Z., Moch, H., Chen, Y.-T., Busam, K.J., et al. (2007). NY-BR-1 is a differentiation antigen of the mammary gland. Appl. Immunohistochem. Mol. Morphol. *15*, 77–83.

Jaye, D.L., Bray, R.A., Gebel, H.M., Harris, W.A.C., and Waller, E.K. (2012). Translational applications of flow cytometry in clinical practice. J. Immunol. *188*, 4715–4719.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Kaplan, D.H., Shankaran, V., Dighe, A.S., Stockert, E., Aguet, M., Old, L.J., and Schreiber, R.D. (1998). Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. Proc. Natl. Acad. Sci. U.S.A. *95*, 7556–7561.

Katzenellenbogen, J.A., O'Malley, B.W., and Katzenellenbogen, B.S. (1996). Tripartite steroid hormone receptor pharmacology: interaction with multiple effector sites as a basis for the cell- and promoter-specific action of these hormones. Mol. Endocrinol. *10*, 119–131.

Kayser, S., Watermann, I., Rentzsch, C., Weinschenk, T., Wallwiener, D., and Gückel, B. (2003). Tumor-associated antigen profiling in breast and ovarian cancer: mRNA, protein or T cell recognition? J. Cancer Res. Clin. Oncol. *129*, 397–409.

Kochenderfer, J.N., Dudley, M.E., Feldman, S.A., Wilson, W.H., Spaner, D.E., Maric, I., Stetler-Stevenson, M., Phan, G.Q., Hughes, M.S., Sherry, R.M., et al. (2012). B-cell depletion and remissions of malignancy along with cytokine-associated toxicity in a clinical trial of anti-CD19 chimeric-antigen-receptor-transduced T cells. Blood *119*, 2709–2720.

Koenig, T., Menze, B.H., Kirchner, M., Monigatti, F., Parker, K.C., Patterson, T., Steen, J.J., Hamprecht, F.A., and Steen, H. (2008). Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. J. Proteome Res. *7*, 3708–3717.

Konno, H., Yamamoto, M., and Ohta, M. (2010). Recent concepts of antiangiogenic therapy. Surg. Today *40*, 494–500.

Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. Cell *127*, 1041–1055.

Kouros-Mehr, H., Kim, J., Bechis, S.K., and Werb, Z. (2008). GATA-3 and the regulation of the mammary luminal cell fate. Current Opinion in Cell Biology *20*, 164–170.

Koutras, A.K., and Evans, T.R.J. (2008). The epidermal growth factor receptor family in breast cancer. Onco Targets Ther *1*, 5–19.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Kumar, B. (2014). Breast Cancer Genomics. In Omics Approaches in Breast Cancer, D. Barh, ed. (New Delhi: Springer India), pp. 53–103.

Kushner, P.J., Agard, D.A., Greene, G.L., Scanlan, T.S., Shiau, A.K., Uht, R.M., and Webb, P. (2000). Estrogen receptor pathways to AP-1. J. Steroid Biochem. Mol. Biol. *74*, 311–317.

Kushwah, R., and Hu, J. (2011). Complexity of dendritic cell subsets and their function in the host immune system. Immunology *133*, 409–419.

Kwiatkowska-Borowczyk, E.P., Gąbka-Buszek, A., Jankowski, J., and Mackiewicz, A. (2015). Review Immunotargeting of cancer stem cells. Współczesna Onkologia *1A*, 52–59.

Laganière, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguère, V. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc. Natl. Acad. Sci. U.S.A. *102*, 11651–11656.

Lehninger, A.L. (2005). Lehninger principles of biochemistry (New York: W.H. Freeman).

Lewandowska, M.A. (2013). The missing puzzle piece: splicing mutations. Int J Clin Exp Pathol *6*, 2675–2682.

Li, G., Pan, T., Guo, D., and Li, L.-C. (2014). Regulatory Variants and Disease: The E-Cadherin −160C/A SNP as an Example. Molecular Biology International *2014*, 1–9.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.-L., Gyorki, D.E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

Liyanage, U.K., Moore, T.T., Joo, H.-G., Tanaka, Y., Herrmann, V., Doherty, G., Drebin, J.A., Strasberg, S.M., Eberlein, T.J., Goedegebuure, P.S., et al. (2002). Prevalence of regulatory T cells is increased in peripheral blood and tumor microenvironment of patients with pancreas or breast adenocarcinoma. J. Immunol. *169*, 2756–2761.

Lodish, H. (2008). Molecular Cell Biology.

Macias, H., and Hinck, L. (2012). Mammary gland development. Wiley Interdiscip Rev Dev Biol *1*, 533–557.

MacKie, R.M., Reid, R., and Junor, B. (2003). Fatal melanoma transferred in a donated kidney 16 years after melanoma surgery. N. Engl. J. Med. *348*, 567–568.

Makarem, M., Spike, B.T., Dravis, C., Kannan, N., Wahl, G.M., and Eaves, C.J. (2013). Stem cells and the developing mammary gland. J Mammary Gland Biol Neoplasia *18*, 209–219.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Marangoni, E., Lecomte, N., Durand, L., de Pinieux, G., Decaudin, D., Chomienne, C., Smadja-Joffe, F., and Poupon, M.-F. (2009). CD44 targeting reduces tumour growth and prevents post-chemotherapy relapse of human breast cancers xenografts. Br. J. Cancer *100*, 918–922.

McGilvray, R.W., Eagle, R.A., Watson, N.F.S., Al-Attar, A., Ball, G., Jafferji, I., Trowsdale, J., and Durrant, L.G. (2009). NKG2D ligand expression in human colorectal cancer reveals associations with prognosis and evidence for immunoediting. Clin. Cancer Res. *15*, 6993–7002.

Medina, D. (2005). Mammary developmental fate and breast cancer risk. Endocrine Related Cancer *12*, 483–495.

Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M.S., Ghosh, D., Chinnaiyan, A.M., and Kleer, C.G. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. Cancer Res. *65*, 11259–11264.

Meier-Abt, F., Milani, E., Roloff, T., Brinkhaus, H., Duss, S., Meyer, D.S., Klebba, I., Balwierz, P.J., van Nimwegen, E., and Bentires-Alj, M. (2013). Parity induces differentiation and reduces Wnt/Notch signaling ratio and proliferation potential of basal stem/progenitor cells isolated from mouse mammary epithelium. Breast Cancer Res. *15*, R36.

Miller, E., Lee, H.J., Lulla, A., Hernandez, L., Gokare, P., and Lim, B. (2014). Current treatment of early breast cancer: adjuvant and neoadjuvant therapy. F1000Res *3*, 198.

Molineris, I., Schiavone, D., Rosa, F., Matullo, G., Poli, V., and Provero, P. (2013). Identification of functional cis-regulatory polymorphisms in the human genome. Hum. Mutat. *34*, 735–742.

Mongan, N., and Gudas, L. (2005). Valproic acid, in combination with all-trans retinoic acid and 5-aza-2'-deoxycytidine, restores expression of silenced RARbeta2 in breast cancer cells. (Molecular Cancer Therapeutics),.

Mougiakakos, D., Choudhury, A., Lladser, A., Kiessling, R., and Johansson, C.C. (2010). Regulatory T cells in cancer. Adv. Cancer Res. *107*, 57–117.

Myc, L.A., Gamian, A., and Myc, A. (2011). Cancer vaccines. Any future? Arch. Immunol. Ther. Exp. (Warsz.) *59*, 249–259.

Narvaez, C.J., Matthews, D., LaPorta, E., Simmons, K.M., Beaudin, S., and Welsh, J. (2014). The impact of vitamin D in breast cancer: genomics, pathways, metabolism. Front Physiol *5*, 213.

Nass, S.J., Herman, J.G., Gabrielson, E., Iversen, P.W., Parl, F.F., Davidson, N.E., and Graff, J.R. (2000). Aberrant methylation of the estrogen receptor and E-cadherin 5' CpG islands increases with malignant progression in human breast cancer. Cancer Res. *60*, 4346–4348.

Ngounou Wetie, A.G., Sokolowska, I., Woods, A.G., Roy, U., Deinhardt, K., and Darie, C.C. (2014). Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. Cell. Mol. Life Sci. *71*, 205–228.

Niederkorn, J.Y. (2009). Immune escape mechanisms of intraocular tumors. Prog Retin Eye Res *28*, 329–347.

Nilsson, S., Mäkelä, S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G., Enmark, E., Pettersson, K., Warner, M., and Gustafsson, J.A. (2001). Mechanisms of estrogen action. Physiol. Rev. *81*, 1535–1565.

No, J.H., Kim, M.-K., Jeon, Y.-T., Kim, Y.-B., and Song, Y.-S. (2011). Human papillomavirus vaccine: widening the scope for cancer prevention. Mol. Carcinog. *50*, 244–253.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.-L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., et al. (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes & Development *22*, 581–586.

Oakes, S.R., Gallego-Ortega, D., and Ormandy, C.J. (2014). The mammary cellular hierarchy and breast cancer. Cell. Mol. Life Sci. *71*, 4301–4324.

Olayioye, M.A. (2001). Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. Breast Cancer Res. *3*, 385–389.

Paech, K., Webb, P., Kuiper, G.G., Nilsson, S., Gustafsson, J., Kushner, P.J., and Scanlan, T.S. (1997). Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. Science *277*, 1508–1510.

Pal, B., Bouras, T., Shi, W., Vaillant, F., Sheridan, J.M., Fu, N., Breslin, K., Jiang, K., Ritchie, M.E., Young, M., et al. (2013). Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. Cell Rep *3*, 411–426.

Parham, P. (2009). The Immune System (Garland Science).

Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M.A., and Daidone, M.G. (2005). Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res. *65*, 5506–5511.

Prince, H.M., Bishton, M.J., and Harrison, S.J. (2009). Clinical studies of histone deacetylase inhibitors. Clin. Cancer Res. *15*, 3958–3969.

Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky, P.V., Jackson, R.B., and Campbell, N.A. (2014). Campbell biology (Boston: Pearson).

Reid, G., Métivier, R., Lin, C.-Y., Denger, S., Ibberson, D., Ivacevic, T., Brand, H., Benes, V., Liu, E.T., and Gannon, F. (2005). Multiple mechanisms induce transcriptional silencing of a subset of genes, including oestrogen receptor alpha, in response to deacetylase inhibition by valproic acid and trichostatin A. Oncogene *24*, 4894–4907.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Rini, B. (2014). Future approaches in immunotherapy. Semin. Oncol. *41 Suppl 5*, S30–S40.

Rosorius, O., Heger, P., Stelz, G., Hirschmann, N., Hauber, J., and Stauber, R.H. (1999). Direct observation of nucleocytoplasmic transport by microinjection of GFP-tagged proteins in living cells. BioTechniques *27*, 350–355.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Roy, A.K., Lavrovsky, Y., Song, C.S., Chen, S., Jung, M.H., Velu, N.K., Bi, B.Y., and Chatterjee, B. (1999). Regulation of androgen action. Vitam. Horm. *55*, 309–352.

Rueckert, S., Ruehl, I., Kahlert, S., Konecny, G., and Untch, M. (2005). A monoclonal antibody as an effective therapeutic agent in breast cancer: trastuzumab. Expert Opin Biol Ther *5*, 853–866.

Santagata, S., Thakkar, A., Ergonul, A., Wang, B., Woo, T., Hu, R., Harrell, J.C., McNamara, G., Schwede, M., Culhane, A.C., et al. (2014). Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J. Clin. Invest. *124*, 859–870.

Sartorelli, V., Puri, P.L., Hamamori, Y., Ogryzko, V., Chung, G., Nakatani, Y., Wang, J.Y., and Kedes, L. (1999). Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. Mol. Cell *4*, 725–734.

Sasakawa, Y., Naoe, Y., Inoue, T., Sasakawa, T., Matsuo, M., Manda, T., and Mutoh, S. (2003). Effects of FK228, a novel histone deacetylase inhibitor, on tumor growth and expression of p21 and c-myc genes in vivo. Cancer Lett. *195*, 161–168.

Saville, B., Wormke, M., Wang, F., Nguyen, T., Enmark, E., Kuiper, G., Gustafsson, J.A., and Safe, S. (2000). Ligand-, cell-, and estrogen receptor subtype (alpha/beta)-dependent activation at GC-rich (Sp1) promoter elements. J. Biol. Chem. *275*, 5379–5387.

Schefe, J.H., Lehmann, K.E., Buschmann, I.R., Unger, T., and Funke-Kaiser, H. (2006). Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. J. Mol. Med. *84*, 901–910.

Scott, G.K., Marden, C., Xu, F., Kirk, L., and Benz, C.C. (2002). Transcriptional repression of ErbB2 by histone deacetylase inhibitors detected by a genomically integrated ErbB2 promoter-reporting cell screen. Mol. Cancer Ther. *1*, 385–392.

Seil, I. (2005). Molekularbiologische Charakterisierung und Expressionsanalyse des Brust Tumorantigens NY-BR-1.

Seil, I., Frei, C., Sültmann, H., Knauer, S.K., Engels, K., Jäger, E., Zatloukal, K., Pfreundschuh, M., Knuth, A., Tseng-Chen, Y., et al. (2007). The differentiation antigen NY-BR-1 is a potential target for antibody-based therapies in breast cancer. Int. J. Cancer *120*, 2635–2642.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.-L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Sorensen, B.S., Wu, L., Wei, W., Tsai, J., Weber, B., Nexo, E., and Meldgaard, P. (2014). Monitoring of epidermal growth factor receptor tyrosine kinase inhibitor-sensitizing and resistance mutations in the plasma DNA of patients with advanced non-small cell lung cancer during treatment with erlotinib. Cancer *120*, 3896–3901.

Stern, D.F. (2003). ErbBs in mammary development. Exp. Cell Res. *284*, 89–98.

Stewart, B., and Wild, C. (2014). World Cancer Report 2014.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Stratton, M.R. (2011). Exploring the genomes of cancer cells: progress and promise. Science *331*, 1553–1558.

Tanos, T., Sflomos, G., Echeverria, P.C., Ayyanan, A., Gutierrez, M., Delaloye, J.-F., Raffoul, W., Fiche, M., Dougall, W., Schneider, P., et al. (2013). Progesterone/RANKL is a major regulatory axis in the human breast. Sci Transl Med *5*, 182ra55.

Theurillat, J.-P., Zürrer-Härdi, U., Varga, Z., Storz, M., Probst-Hensch, N.M., Seifert, B., Fehr, M.K., Fink, D., Ferrone, S., Pestalozzi, B., et al. (2007). NY-BR-1 protein expression in breast carcinoma: a mammary gland differentiation antigen as target for cancer immunotherapy. Cancer Immunol. Immunother. *56*, 1723–1731.

Theurillat, J.-P., Zürrer-Härdi, U., Varga, Z., Barghorn, A., Saller, E., Frei, C., Storz, M., Behnke, S., Seifert, B., Fehr, M., et al. (2008). Distinct expression patterns of the immunogenic differentiation antigen NY-BR-1 in normal breast, testis and their malignant counterparts. Int. J. Cancer *122*, 1585–1591.

Tzortzatou-Stathopoulou, F. (1994). History of pediatric hematology and oncology in Greece. Pediatr Hematol Oncol *11*, 13–25.

Varga, Z., Theurillat, J.-P., Filonenko, V., Sasse, B., Odermatt, B., Jungbluth, A.A., Chen, Y.-T., Old, L.J., Knuth, A., Jäger, D., et al. (2006). Preferential nuclear and cytoplasmic NY-BR-1 protein expression in primary breast cancer and lymph node metastases. Clin. Cancer Res. *12*, 2745–2751.

Varley, K.E., Gertz, J., Roberts, B.S., Davis, N.S., Bowling, K.M., Kirby, M.K., Nesmith, A.S., Oliver, P.G., Grizzle, W.E., Forero, A., et al. (2014). Recurrent read-through fusion transcripts in breast cancer. Breast Cancer Res. Treat. *146*, 287–297.

Velasco-Velázquez, M.A., Homsi, N., De La Fuente, M., and Pestell, R.G. (2012). Breast cancer stem cells. Int. J. Biochem. Cell Biol. *44*, 573–577.

Vicente, J.J., and Wordeman, L. (2015). Mitosis, microtubule dynamics and the evolution of kinesins. Exp. Cell Res.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. *28*, 1143–1158.

Wallace, P.K., Tario, J.D., Fisher, J.L., Wallace, S.S., Ernstoff, M.S., and Muirhead, K.A. (2008). Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution. Cytometry A *73*, 1019–1034.

Wallwiener, C.W., Wallwiener, M., Kurth, R.R., Röhm, C., Neubauer, H., Banys, M.J., Staebler, A., Schönfisch, B., Meuer, S.C., Giese, T., et al. (2011). Molecular detection of breast cancer metastasis in sentinel lymph nodes by reverse transcriptase polymerase chain reaction (RT-PCR): identifying, evaluating and establishing multi-marker panels. Breast Cancer Research and Treatment *130*, 833–844.

Wang, C., Fu, M., Angeletti, R.H., Siconolfi-Baez, L., Reutens, A.T., Albanese, C., Lisanti, M.P., Katzenellenbogen, B.S., Kato, S., Hopp, T., et al. (2001). Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity. J. Biol. Chem. *276*, 18375–18383.

Wang, X., Huang, T., Zhao, Y., Zheng, Q., Thompson, R.C., Bu, G., Zhang, Y., Hong, W., and Xu, H. (2014). Sorting Nexin 27 Regulates Aβ Production through Modulating γ-Secretase Activity. Cell Reports *9*, 1023–1033.

Weber, J.S. (2014). Current perspectives on immunotherapy. Semin. Oncol. *41 Suppl 5*, S14–S29.

Weinberg, R.A. (2014). The biology of cancer (Taylor & Francis Ltd.).

Widschwendter, M., and Jones, P.A. (2002). DNA methylation and breast carcinogenesis. Oncogene *21*, 5462–5482.

Woo, E.Y., Chu, C.S., Goletz, T.J., Schlienger, K., Yeh, H., Coukos, G., Rubin, S.C., Kaiser, L.R., and June, C.H. (2001). Regulatory CD4(+)CD25(+) T cells in tumors from patients with early-stage non-small cell lung cancer and late-stage ovarian cancer. Cancer Res. *61*, 4766–4772.

Worby, C.A., and Dixon, J.E. (2002). Sorting out the cellular functions of sorting nexins. Nat. Rev. Mol. Cell Biol. *3*, 919–931.

Yamamoto-Ibusuki, M., Yamamoto, Y., Fujiwara, S., Sueta, A., Yamamoto, S., Hayashi, M., Tomiguchi, M., Takeshita, T., and Iwase, H. (2014). C6ORF97-ESR1 breast cancer susceptibility locus: influence on progression and survival in breast cancer patients. Eur. J. Hum. Genet.

Yang, X., Ferguson, A.T., Nass, S.J., Phillips, D.L., Butash, K.A., Wang, S.M., Herman, J.G., and Davidson, N.E. (2000). Transcriptional activation of estrogen receptor alpha in human breast cancer cells by histone deacetylase inhibition. Cancer Res. *60*, 6890–6894.

Yao, Y.L., Yang, W.M., and Seto, E. (2001). Regulation of transcription factor YY1 by acetylation and deacetylation. Mol. Cell. Biol. *21*, 5979–5991.

Yersal, O., and Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol *5*, 412–424.

Zinser, G., Packman, K., and Welsh, J. (2002). Vitamin D(3) receptor ablation alters mammary gland morphogenesis. Development *129*, 3067–3076.

8 Manuscript

In silico SNP analysis of the breast cancer antigen NY-BR-1

Julia Bitzer[1], Zeynep Kosaloglu[2], Niels Halama[1], Zhiqin Huang[3], Marc Zapatka[3], Peter Lichter[3], Dirk Jäger[1,2] and Inka Zörnig[1]

[1] Department of Medical Oncology, National Center for Tumor Diseases (NCT) and University Hospital Heidelberg, Heidelberg, Germany

[2] Clinical Cooperation Unit "Applied Tumor Immunity", National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany

[3] Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Abstract

Breast cancer is one of the most common malignancies with increasing incidences every year and a leading cause of death among women. Although early stage breast cancer can be effectively treated, there are limited numbers of treatment options available for patients with advanced and metastatic disease. The novel breast cancer associated antigen NY-BR-1 was identified by SEREX analysis and is expressed in the majority (>70 %) of breast tumors as well as metastases, in normal breast tissue, in testis and occasionally in prostate tissue. The biological function and regulation of NY-BR-1 is up to date unknown. Therefore, we performed an in silico analysis on the genetic variations of the NY-BR-1 gene using data available in public SNP databases and several computer programs.

Over 2800 SNPs are recorded in the dbSNP and NHLBI ESP databases for the NY-BR-1 gene. Of these, 65 (2.07 %) are synonymous SNPs, 191 (6.09 %) are non-synoymous SNPs, and 2430 (77.48 %) are noncoding intronic SNPs. We analyzed the 191 non-synonymous SNPs using the tools SIFT, Polyphen and Provean to find possible functional SNPs. As a result, 69 non-synoymous SNPs were predicted to be damaging by at least two, and 16 SNPs were predicted as damaging by all three of the used tools. The SNPs rs200639888, rs367841401 and rs377750885 were categorized as highly damaging by all three tools. Eight damaging SNPs are located in the ankyrin repeat domain (ANK), a domain known for its frequent involvement in protein-protein interactions. Considering these results we expect to gain more insights into the variations of the NY-BR-1 gene and their possible impact on

giving rise to splice variants and therefore influence the function of NY-BR-1 in healthy tissue as well as in breast cancer.

Introduction

Breast cancer is one of the most common malignancies and a leading cause of death among women. Although early stage breast cancer can be effectively treated, there are limited numbers of treatment options available for patients with advanced and metastatic disease. Therefore new targets and strategies need to be developed. A novel breast cancer differentiation antigen, designated as New York-Breast-1 (NY-BR-1), was identified by a serological cloning strategy (SEREX) [1, 2] and could be a possible target for immunotherapy for breast cancer patients [3]. NY-BR-1, also known as ANKRD30A, is located on chromosome 10p11-p12. There are several transcripts existing, which contain between 36 and 42 exons. Although computational analyses have identified NY-BR-1 as being a potential transcription factor, the functional aspects of this 158.9 kDa protein are still unknown. NY-BR-1 protein was shown to be expressed in normal breast epithelia cells and in a majority of primary breast cancers [4, 5], while NY-BR-1 mRNA was detected predominantly in breast cancers [6, 7]. NY-BR-1 is overexpressed in over 70 % of primary breast tumors and metastases [1] and additional details on the involvement of NY-BR-1 in breast cancer will lead to a better understanding of the underlying processes.

Information on genetic variation can provide a valuable insight into the functional range and critical regions of a gene. In fact, gene function cannot be fully understood without awareness of the potential variability within a gene [8]. To further understand the biological function and regulation of NY-BR-1 and its potential for therapeutic approaches, we performed an *in silico* analysis on the genetic variations of the NY-BR-1 gene.

Genetic variation may be diverse in nature, ranging from single nucleotide polymorphisms (SNPs), tandem repeats, small insertions or deletions (indels) to large copy number variations. Among these, SNPs are the most common form of human variation and it has been estimated that one SNP exists every 290 base-pairs in the human genome [9]. Evidences show that through SNPs a wide range of human diseases such as cancer or autoimmunity can be triggered. SNPs also might affect the pharmacokinetics and pharmacodynamics of certain drugs in cancer therapy [10]. The transcriptional regulation of a protein, its structure and its function can be affected by a single base substitution, deletion or insertion. Two groups of SNPs are known: synonymous (sSNP) and non- synonymous SNPs (nsSNP). The latter results in changes of the translated amino acid sequence. SNPs can be located in introns

(affecting gene expression) and exons (altering protein structure or function) of a gene, and also in untranslated regions (UTRs) resulting in an altered stability of the mRNA [11].

The growing number of known SNPs combined with the growing functional annotation of the human genome enabled the analysis of links between genetic and phenotypic variation in a genome-wide manner. Genome-wide association studies (GWAS) aim to detect links between genetic variants (i.e. SNPs) and the appearance of certain phenotypes (i.e. complex diseases). A number of studies have shown associations between one or few SNPs and complex diseases, but until today it is not entirely clear how much impact SNPs have on certain traits in different populations.

With the steadily increasing number of known human nsSNPs, there is also growing interest in identification of the subset that may affect protein function. Various types of features can be used to predict the functional impact of nsSNPs: physical and chemical properties of the affected amino acids, structural properties of the encoded protein, and evolutionary properties, which can be inferred from sequence alignments of homologous proteins [12]. SIFT (Sorting Intolerant from Tolerant) [13], PROVEAN (Protein Variation Effect Analyzer) [14] and PolyPhen-2 (Polymorphism Phenotyping v2) [15] are computational prediction methods which take several of these properties into account and calculate a score to predict whether a given nsSNP has a functional impact. We obtained all SNPs for the NY-BR-1 gene and investigated the nsSNPs for their functional impact by using these three prediction tools. We identified a small number of nsSNPs which seem to affect the protein function of NY-BR-1. Additionally, we used in house sequencing data to analyze whether certain SNPs are enriched in breast cancer patients.

Materials and Methods

SNP Mining

dbSNP is hosted by the National Center for Biotechnology Information (NCBI) and is the largest repository of SNP data with over 140 million submitted variations [16].

Another source of variation data is provided by the "The National Heart, Lung and Blood Institute" (NHLBI). With the aim of discovering novel genes and mechanisms contributing to heart, lung and blood disorders, the NHLBI started the Exome Sequencing Project (ESP) and a large and well-phenotyped population with over 200,000 individuals was assembled. The protein coding regions of each individual genome (i.e. exome) is sequenced and the variation data is made publicly available [17].

The Ensembl Variation database incorporates variation data from several sources including dbSNP and NHLBI ESP. We used the web interface MartWizard (http://central.biomart.org/martwizard/) of the BioMart Central Portal which offers access and crosslinks a wide array of biological databases.

The Ensembl transcript ID ENST00000611781 of the ANKRD30A gene was used to retrieve all available germline variations together with the corresponding genomic coordinates, the variant descriptions, and the validation status. Using the variant descriptions, we filtered coding non-synonymous SNPs (nsSNPs), coding synonymous SNPs (sSNPs) and intronic SNPs.

Additionally, we were provided with exome-sequencing data of 55 breast cancer patients from an in-house sequencing project. We also analyzed this dataset and looked for SNPs which are recorded in dbSNP.

Prediction of the functional impact of coding nsSNPs using SIFT

The prediction tool SIFT evaluates the functional impact of SNPs based on sequence homology. The prediction is based on the degree of conservation of each amino acid residue of the query sequence. To assess the degree of conservation, SIFT compiles a dataset of functionally related protein sequences by searching the protein databases UniProt and TrEMBL using the PSI-BLAST algorithm and builds an alignment of the found sequences and the query sequence. In the second step a normalized probability for each substitution at each position of the alignment is calculated and is then recorded in a scaled probability matrix. This scaled probability is also called the SIFT score and a substitution is considered to be tolerated if the score is greater than 0.05; those less than 0.05 are predicted to be

deleterious. The SIFT approach assumes that a highly conserved position is intolerant to most substitutions, whereas a poorly conserved position can tolerate most substitutions.

Prediction of the functional impact of coding nsSNPs using PROVEAN

The tool PROVEAN also uses an alignment approach to assesses the functional impact of SNPs. PROVEAN consists of two main steps. In the first step, a set of homologous and distantly related sequences from the NCBI NR protein database is collected using BLASTP. To remove redundancy, the collected sequences are clustered, based on a sequence identity of 80 %. A so called supporting set of sequences is assembled by adding sequences from clusters most similar to the query sequence, until a sufficient number of clusters is reached in the supporting set. In the second step, for each sequence in the supporting sequence set, a delta score is computed using the BLOSUM62 substitution matrix. For each cluster, an average delta score is computed, and the averaged delta scores are again averaged among all clusters. This unbiased averaged delta score is the final PROVEAN score.

The PROVEAN approach assumes that a variation, which reduces similarity of protein A to the homologous or distantly related protein B, is more likely to cause a damaging effect. Thus, the impact of a variation on protein function can be measured as the change in alignment score, the delta score. Low delta scores are interpreted as variations leading to a deleterious effect on protein function, while high delta scores are interpreted as variations with neutral effect.

The tools SIFT and PROVEAN are available online at http://provean.jcvi.org/. On the website, we used the tool PROVEAN Human Genome Variants, which provides PROVEAN and SIFT predictions for a list of human genome variants. We submitted the list of genomic coordinates and variants of our filtered 191 nsSNPs, and chose the default threshold of delta score <= -2.5 to detect deleterious variations.

Prediction of the functional impact of coding nsSNPs using PolyPhen-2

PolyPhen-2 combines information on sequence features, multiple alignments with homologous proteins, and structural parameters to predict the impact of a SNP on protein function.

For sequence-based assessment, PolyPhen-2 tries to identify the query as an entry in the UniProtKB/Swiss-Prot database. Using the feature table of the corresponding entry, PolyPhen-2 checks if a given SNP occurs at functional relevant site, e.g. if the SNP lies within a transmembrane, signal peptide, or binding region.

Similar to SIFT, PolyPhen-2 also assesses the degree of conversation of the position where the SNP occurs by utilizing a multiple sequence alignment of homologous sequences. For each variant PolyPhen-2 calculates a position-specific independent counts (PSIC) score. The PSIC score difference between the two variants describes the impact of a particular amino acid substitution: the higher the PSIC score difference, the higher functional impact the substitution is likely to have.

A BLAST query of the query sequence against protein structure databases is carried out to identify corresponding 3D protein structures. If corresponding structures are found, they are used to assess, whether the SNP is likely to destroy the hydrophobic core, interactions with ligands or other important features of the protein.

Finally, all parameters are taken together and empirical prediction rules are applied to make the final decision, whether the SNP is damaging or benign.

PolyPhen-2 is available online at http://genetics.bwh.harvard.edu/pph2/. We used the option ‘Batch query’ and submitted the list of genomic coordinates and variants of our filtered 191 nsSNPs.

Results

SNP Mining

In the Ensembl BioMart database 2898 SNPs are recorded for the ANKRD30A transcript ENST00000611781. 2880 of these were imported from dbSNP and 18 from NHLBI ESP. 1832 SNPs have been validated by in dependent submissions or frequency/genotype data. However, the clinical significance has not been determined yet for any of the SNPs.

Out of all 2898 SNPs, 65 (2.07 %) were sSNPs, 191 (6.09 %) were nsSNPs, and 2430 (77.48 %) occurred in intronic regions (Figure 1). 40 of the downloaded SNPs are annotated as splice region variants in dbSNP. We selected nsSNPs for our investigation.

Deleterious nsSNPs predicted by SIFT

Among the 191 analyzed nsSNPs, 79 nsSNPs were identified to be damaging with a tolerance index score >= 0.5. Ten nsSNPS showed a highly damaging tolerance index score of 0.00, namely rs200639888, rs372199195, rs144539033, rs369532435, rs199571878, rs376821949, rs267602482, rs201234943, rs367841401, and rs377750885. Nine nsSNPs had a tolerance index score of 0.001, nine nsSNPs had a score of 0.002, and five had a score of 0.003. The remaining nsSNPs contained tolerance index scores varying between 0.004 and 0.048.

Damaging nsSNPs predicted by PROVEAN

28 nsSNPs out of the analyzed 191 nsSNPs were predicted to be deleterious with a delta score of <= -2.5. 10 nsSNPs showed a highly deleterious score of < - 4.00: rs200639888 (-5.962), rs61737412(-5.030), rs201943652(-4.758), rs189195791(-6.263), rs367841401 (-4.465), rs185294248(-4.366), rs374753521 (-4.184), rs371981371 (-4.603), rs377750885(-5.87), and rs201764363(-4.025).

20 nsSNPs were predicted as damaging variations by SIFT and PROVEAN. rs200639888, rs367841401, and rs377750885 were predicted to be highly damaging by SIFT with a tolerance index score of 0.00 and are also predicted to be highly deleterious by PROVEAN with delta scores of -5.962 and -4.465, and -5.87 respectively.

Damaging nsSNPs predicted by PolyPhen

Out of the 171 nsSNPs submitted to the PolyPhen-2 server, 102 nsSNPs were considered to be damaging: 44 nsSNPs were predicted to be ‘probably damaging’ with an PSIC score of 2.00 or more, and 58 nsSNPs were predicted to be ‘possibly damaging’ with an PSIC score of 1.40-1.90. The remaining 89 nsSNPs were predicted to be benign.

64 of the nsSNPs which were predicted to be damaging by SIFT, were also predicted damaging by PolyPhen. rs200639888, rs369532435, rs267602482, rs201234943, and rs377750885 were among the nsSNPs predicted to be highly damaging by SIFT with a tolerance index score of 0.00. These five nsSNPs also have high PSIC scores predicted by PolyPhen (2.439, 2.746, 2.23, 2.373, and 2.46 respectively).

19 nsSNPs were predicted to be damaging by Provean and PolyPhen, and 16 nsSNPs were predicted to be damaging by all three of the used tools (Figure 2). The nsSNPs rs200639888, rs367841401, and rs377750885 were predicted to be highly damaging/deleterious by all three tools.

Damaging nsSNPs predicted by at least two tools

As summarized in Table 1, 69 nsSNPs were predicted to be damaging/deleterious by at least two of the used tools. We selected these 69 nsSNPs to perform a more detailed analysis.

Analysis of the spectrum of nsSNPs on the nucleotide level showed a conserved profile with A>T/T>A transitions being the most frequent changes (Figure 3a). We further analyzed the spectrum of amino acid changes and found that hydrophile>hydrophile transitions occur most frequently and hydrophobe>hydrophobe transitions second frequently (Figure 3b).

The nsSNPs rs200639888 and rs367841401, which were predicted to be highly damaging by all three tools, have an amino acid change from leucine to proline which are hydrophobe amino acids. The third damaging nsSNP predicted by all three tools, rs377750885, has a change from glutamic acid (hydrophile) to (hydrophobe) valine.

Discussion

Information on genetic variation can provide a valuable insight into the functional range and critical regions of a gene. SNPs are the most common form of genetic variations and a link between SNPs and complex diseases have been reported for a number of cases. The BRCA-1 gene for example and some of its interaction partners are associated with breast cancer. SNPs in these genes are not just involved in the onset of a disease but they can promote also disease progression and outcome [18, 19]. Here, we systematically analyzed SNPs in the NY-BR-1 gene to identify those SNPs which can modify the functional properties of the protein.

In the Ensembl BioMart database 2898 SNPs are recorded for the NY-BR-1 transcript ENST00000611781. Out of these, 191 (6.01%) were nonsynonymous SNPs (nsSNPs), i.e. polymorphisms which translate into an altered amino acid sequence. As these types of SNPs are most likely to have an effect on protein function, we chose to analyze only them further.

Computational approaches use various types of features to predict the functional impact of nsSNPs: physical and chemical properties of the affected amino acids, structural properties of the encoded protein, and evolutionary properties, which can be inferred from sequence alignments of homologous proteins. We chose three state-of-the-art computational tools which can predict the effects of amino acid substitutions on protein function: SIFT, Provean and PolyPhen2.

191 nsSNPs were analyzed and the results varied between the used tools: SIFT predicted 79 damaging nsSNPs, Provean 28 nsSNPs, and PolyPhen2 102 nsSNPs. 16 nsSNPs were predicted damaging by all three tools, and a total of 69 nsSNPs were predicted damaging by at least two of the used tools. SIFT and PolyPhen2 have the biggest overlap with 63 common predictions. This may be due to the common step of assessing the degree of conversation by utilizing a multiple sequence alignment of homologous sequences. 36 damaging nsSNPs were only predicted by PolyPhen2 because PolyPhen2 is the only tool that takes functional relevant sites into account. The location of the 69 damaging SNPS within the ANKRD30A gene is shown in Figure 4a.

Up to date the structure of NY-BR-1 has not been solved yet and no homology models for the entire protein are available. Thus, we unfortunately could not evaluate the location and effect of the predicted damaging nsSNPs on the protein structure.

In the UniProt database six ankyrin (ANK) repeat motifs are documented for NY-BR-1. The ANK repeat motif is one of the most common protein-protein interaction motifs in nature and occurs in a large number of functionally diverse proteins. The structure of the ANK repeat motif is conserved: each repeat typically consists of 30–34 amino acid residues comprising

two anti-parallel α-helices and a long loop ending in a β-hairpin [20]. Proteins containing the ANK repeat motif are involved in a diverse set of cellular functions, and defects in ANK repeat proteins have been associated with a number of human diseases [21, 22]. Hence, a variation within such a functional domain is likely to have an impact on protein function.

Eight of the 69 damaging nsSNPs in NY-BR-1 are located in an ANK repeat motif: rs190686350, rs200639888, rs61737412, rs267602477, rs201943652, rs200651327, rs369118323, rs373380909. PolyPhen2 predicted all of them as damaging, whereas Provean predicted two, and SIFT one of them as not damaging.

To understand genetic risk factors it is important how SNPs can change protein structure and function because different transcripts can lead to a new structure and function or degradation of the protein. Alternative splicing is one of the key drivers to regulate the transcriptome. There are four major types known of alternative splicing: exon skipping, alternative 3' or 5' splice site, and intron retention. Genome studies revealed that exon skipping occurs on a frequent basis whereas intron retention is observed on a low frequency. It was assumed that intron retentions are derived from non-spliced or partially spliced pre-mRNAs [23]. Intron retention can result in mRNA accumulation and protein expression [24]. The alternative splicing process can be influenced and controlled by SNPs. If SNPs are located at a donor or acceptor site they can affect the splicing machinery. In dbSNP, 39 of the NY-BR-1 SNPs are annotated to be splicing, located at donor or acceptor sites (Figure 4b). These SNPs have the potential to influence the splicing process and new transcripts can be generated.

An unknown fraction of SNPs submitted to the public databases may not be true polymorphisms, but examples of sequencing errors. Therefore it is important to consider the validation status of each SNP. A polymorphism can be validated by independent submissions or frequency/genotype data. In our dataset 1832 out of 2898 SNPs have been validated. Considering the 69 damaging nsSNPs, 16 have not been validated yet. As these nsSNPs seem to have an impact on protein function, validation of them should especially be considered.

We also analyzed in house exome-sequencing data of 55 breast cancer patients, and found 11 SNPs in the NY-BR-1 gene in this patient cohort. Seven SNPs occur in more than 10 patients and three of these (rs61737412, rs41276130, rs1200875) were predicted damaging by at least two of the used tools. This additionally suggests clinical relevance for these three damaging SNPs.

In summary, we have identified 69 damaging nsSNPs within the coding region of the breast cancer associated NY-BR-1 gene. Moreover, we found 39 potential splicing SNPs which can affect the alternative splicing process. Our analysis gives an overview on the SNP landscape

of NYBR1 and now provides the basis to further study the association of SNPs and the molecular breast cancer subtypes “Her2”, “Luminal A/B” and “Triple negative” as well as clinical data, such as treatment response, relapse rate and overall survival.

References

1. Jager, D. et al., *Humoral and cellular immune responses against the breast cancer antigen NY-BR-1: definition of two HLA-A2 restricted peptide epitopes.* Cancer Immun, 2005. **5**: p. 11.
2. Jager, D. et al., *Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library.* Cancer Res, 2001. **61**(5): p. 2055-61.
3. Seil, I. et al., *The differentiation antigen NY-BR-1 is a potential target for antibody-based therapies in breast cancer.* Int J Cancer, 2007. **120**(12): p. 2635-42.
4. Jager, D. et al., *NY-BR-1 is a differentiation antigen of the mammary gland.* Appl Immunohistochem Mol Morphol, 2007. **15**(1): p. 77-83.
5. Varga, Z. et al., *Preferential nuclear and cytoplasmic NY-BR-1 protein expression in primary breast cancer and lymph node metastases.* Clin Cancer Res, 2006. **12**(9): p. 2745-51.
6. Jiang, Y. et al., *Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays.* Oncogene, 2002. **21**(14): p. 2270-82.
7. Nissan, A. et al., *Multimarker RT-PCR assay for the detection of minimal residual disease in sentinel lymph nodes of breast cancer patients.* Br J Cancer, 2006. **94**(5): p. 681-5.
8. Barnes, M.R., *Genetic variation analysis for biomedical researchers: a primer.* Methods Mol Biol, 2010. **628**: p. 1-20.
9. Berriman, M. and A. Pain, *Variety is the spice of eukaryotic life.* Nat Rev Microbiol, 2007. **5**(9): p. 660-1.
10. Wang, J.B. et al., *SNP Web Resources and Their Potential Applications in Personalized Medicine.* Current Drug Metabolism, 2012. **13**(7): p. 978-990.
11. Dickinson, A.M., *Non-HLA genetics and predicting outcome in HSCT.* International Journal of Immunogenetics, 2008. **35**(4-5): p. 375-380.
12. Nakken, S., I. Alseth, and T. Rognes, *Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes.* Neuroscience, 2007. **145**(4): p. 1273-1279.

13. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nature Protocols, 2009. **4**(7): p. 1073-1082.

14. Choi, Y. et al., *Predicting the Functional Effect of Amino Acid Substitutions and Indels.* Plos One, 2012. **7**(10).

15. Adzhubei, I.A. et al., *A method and server for predicting damaging missense mutations.* Nature Methods, 2010. **7**(4): p. 248-249.

16. Sherry, S.T. et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Research, 2001. **29**(1): p. 308-311.

17. Dorschner, M.O. et al., *Actionable, Pathogenic Incidental Findings in 1,000 Participants' Exomes.* American Journal of Human Genetics, 2013. **93**(4): p. 631-640.

18. Alshatwi, A.A. et al., *Identification of Functional SNPs in BARD1 Gene and In Silico Analysis of Damaging SNPs: Based on Data Procured from dbSNP Database.* Plos One, 2012. **7**(10).

19. Johnson, N. et al., *Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility.* Human Molecular Genetics, 2007. **16**(9): p. 1051-1057.

20. Chakrabarty, B. and N. Parekh, *Identifying tandem Ankyrin repeats in protein structures.* BMC Bioinformatics, 2014. **15**(1): p. 6599.

21. Leite, R.C. et al., *Low frequency of ankyrin mutations in hereditary spherocytosis: identification of three novel mutations.* Hum Mutat, 2000. **16**(6): p. 529.

22. Li, J. A. Mahajan, and M.D. Tsai, *Ankyrin repeat: a unique motif mediating protein-protein interactions.* Biochemistry, 2006. **45**(51): p. 15168-78.

23. Galante P.A., et al., *Detection and evaluation of intron retention events in the human transcriptome.* RNA, 2004. **10**(5): p. 757-65.

24. Kralovicova, J. et al., *Optimal antisense target reducing INS intron 1 retention is adjacent to a parallel G quadruplex.* Nucleic Acids Res, 2014. **42**(12): p. 8161-73.

Table 1: Summary of all 69 nsSNPs predicted to be damaging/deleterious by at least two of the used tools.

ID	chromLoc	protLoc	nt_var	prot_var	sift_entry	provean_entry	polyphen_entry	aa_ch	in_ank
rs190686350	37419160	122	G/A	A/T	Damaging(0.012)	Deleterious(-3.27)	probably damaging(0.997)	hydrophobe>hydrophile	Yes
rs200639888	37419170	125	T/C	L/P	Damaging(0)	Deleterious(-5.96)	probably damaging(0.997)	hydrophobe>hydrophobe	Yes
rs61737412	37419218	141	C/T	T/M	Damaging(0.035)	Deleterious(-5.03)	possibly damaging(0.951)	hydrophile>hydrophobe	Yes
rs267602477	37419220	142	G/A	A/T	Damaging(0.016)	Deleterious(-3.41)	probably damaging(0.997)	hydrophobe>hydrophile	Yes
rs201943652	37421175	173	T/C	L/P	Damaging(0.011)	Deleterious(-4.76)	probably damaging(0.995)	hydrophobe>hydrophobe	Yes
rs373997768	37505217	993	A/C	K/T	Damaging(0.003)	Deleterious(-2.76)	probably damaging(0.963)	hydrophile>hydrophile	
rs140013037	37505242	1001	G/C	K/N	Damaging(0.016)	Deleterious(-3.58)	probably damaging(0.963)	hydrophile>hydrophile	
rs200399695	37506718	1060	G/C	R/T	Damaging(0.029)	Deleterious(-2.91)	benign(0.013)	hydrophile>hydrophile	
rs267602485	37507968	1110	G/A	E/K	Damaging(0.021)	Deleterious(-3.34)	probably damaging(0.98)	hydrophile>hydrophile	
rs367841401	37508002	1121	T/C	L/P	Damaging(0)	Deleterious(-4.47)	probably damaging(0.969)	hydrophobe>hydrophobe	
rs185294248	37508038	1133	T/C	F/S	Damaging(0.012)	Deleterious(-4.37)	benign(0.006)	hydrophobe>hydrophile	
rs373377344	37508379	1247	G/A	E/K	Damaging(0.048)	Deleterious(-2.78)	possibly damaging(0.898)	hydrophile>hydrophile	
rs374753521	37508446	1269	A/C	Y/S	Damaging(0.031)	Deleterious(-4.18)	benign(0.347)	hydrophile>hydrophile	
rs199841724	37508538	1300	C/A	H/N	Damaging(0.032)	Deleterious(-3.11)	benign(0.03)	hydrophile>hydrophile	
rs371878855	37508548	1303	A/G	Q/R	Damaging(0.012)	Deleterious(-2.52)	possibly damaging(0.808)	hydrophile>hydrophile	
rs369099906	37508651	1337	A/T	L/F	Damaging(0.001)	Deleterious(-3.22)	probably damaging(0.999)	hydrophobe>hydrophobe	
rs371981371	37508671	1344	C/A	A/D	Damaging(0.003)	Deleterious(-4.6)	probably damaging(0.989)	hydrophobe>hydrophile	
rs200929491	37508788	1383	G/A	R/H	Damaging(0.002)	Deleterious(-3.55)	probably damaging(0.987)	hydrophile>hydrophile	
rs377750885	37508803	1388	A/T	E/V	Damaging(0)	Deleterious(-5.87)	probably damaging(0.997)	hydrophile>hydrophobe	
rs201764363	37508814	1392	G/C	A/P	Damaging(0.002)	Deleterious(-4.03)	probably damaging(0.969)	hydrophobe>hydrophobe	
rs200651327	37418912	105	G/A	E/K	Tolerated(0.081)	Deleterious(-3.37)	probably damaging(0.999)	hydrophile>hydrophile	Yes
rs377744149	37508352	1238	G/A	D/N	Tolerated(0.083)	Deleterious(-3.19)	probably damaging(0.971)	hydrophile>hydrophile	
rs201858051	37508539	1300	A/G	H/R	Tolerated(0.108)	Deleterious(-3.06)	possibly damaging(0.651)	hydrophile>hydrophile	
rs369118323	37422851	209	C/T	L/F	Damaging(0.01)	Neutral(-1.52)	probably damaging(0.993)	hydrophobe>hydrophobe	Yes
rs373380909	37422972	249	G/T	G/V	Damaging(0.003)	Neutral(-2.41)	probably damaging(0.999)	hydrophobe>hydrophobe	Yes
rs202090351	37430699	292	C/A	P/T	Damaging(0.001)	Neutral(-1.91)	probably damaging(0.998)	hydrophobe>hydrophile	
rs377740138	37430720	299	G/A	V/M	Damaging(0.002)	Neutral(-0.39)	possibly damaging(0.845)	hydrophobe>hydrophobe	
rs200845385	37430796	324	C/T	T/I	Damaging(0.002)	Neutral(-0.78)	possibly damaging(0.898)	hydrophile>hydrophobe	
rs372199195	37430803	326	T/G	D/E	Damaging(0)	Neutral(-0.13)	possibly damaging(0.643)	hydrophile>hydrophile	
rs371384886	37430859	345	C/T	A/V	Damaging(0.001)	Neutral(-0.56)	probably damaging(0.997)	hydrophobe>hydrophobe	
rs202098264	37430875	350	C/G	F/L	Damaging(0.003)	Neutral(-1.01)	probably damaging(0.965)	hydrophobe>hydrophobe	
rs372420008	37430922	366	A/G	K/R	Damaging(0.007)	Neutral(-0.62)	possibly damaging(0.956)	hydrophile>hydrophile	
rs374024060	37430943	373	C/T	T/M	Damaging(0.011)	Neutral(-0.76)	probably damaging(0.975)	hydrophile>hydrophobe	
rs144539033	37430978	385	T/C	W/R	Damaging(0)	Neutral(-1.04)	possibly damaging(0.943)	hydrophobe>hydrophile	
rs371443557	37431010	395	T/G	I/M	Damaging(0.004)	Neutral(-0.28)	possibly damaging(0.676)	hydrophobe>hydrophobe	
rs200264724	37431045	407	C/T	T/M	Damaging(0.001)	Neutral(0.55)	probably damaging(0.989)	hydrophile>hydrophobe	
rs369532435	37438591	519	A/T	K/M	Damaging(0)	Neutral(-1.13)	probably damaging(0.996)	hydrophile>hydrophobe	
rs267602481	37438727	532	C/T	S/F	Damaging(0.001)	Neutral(-1.27)	possibly damaging(0.842)	hydrophile>hydrophobe	
rs199571878	37438753	541	A/C	K/Q	Damaging(0)	Neutral(-0.47)	possibly damaging(0.947)	hydrophile>hydrophile	
rs376821949	37438772	547	G/A	R/K	Damaging(0)	Neutral(0.04)	possibly damaging(0.643)	hydrophile>hydrophile	
rs45515098	37440991	550	C/T	P/L	Damaging(0.028)	Neutral(-1.27)	possibly damaging(0.581)	hydrophobe>hydrophobe	
rs377410013	37440994	551	T/C	M/T	Damaging(0.045)	Neutral(-0.42)	possibly damaging(0.717)	hydrophobe>hydrophile	
rs267602482	37441009	556	C/T	S/F	Damaging(0)	Neutral(-1.66)	probably damaging(0.99)	hydrophile>hydrophobe	
rs374037740	37441038	566	T/G	W/G	Damaging(0.009)	Neutral(-1.82)	possibly damaging(0.826)	hydrophobe>hydrophobe	
rs372878721	37442530	580	G/A	V/M	Damaging(0.013)	Neutral(-0.77)	probably damaging(0.976)	hydrophobe>hydrophobe	
rs368660392	37442552	587	A/G	H/R	Damaging(0.003)	Neutral(-1.4)	possibly damaging(0.932)	hydrophile>hydrophile	
rs201669885	37447325	602	C/G	P/A	Damaging(0.012)	Neutral(-1.66)	possibly damaging(0.924)	hydrophobe>hydrophobe	
rs201976592	37447446	611	C/A	T/N	Damaging(0.002)	Neutral(-1.14)	possibly damaging(0.851)	hydrophile>hydrophile	
rs113525905	37447451	613	G/A	G/R	Damaging(0.013)	Neutral(-1.25)	probably damaging(0.998)	hydrophobe>hydrophile	
rs201234943	37447491	626	A/T	K/M	Damaging(0)	Neutral(-1.29)	probably damaging(0.98)	hydrophile>hydrophobe	

rs371253665	37451583	636	C/T	P/L	Damaging(0.001)	Neutral(-0.77)	probably damaging(0.994)	hydrophobe>hydrophobe	
rs199874591	37451705	644	C/A	P/H	Damaging(0.001)	Neutral(-1.11)	probably damaging(0.997)	hydrophobe>hydrophile	
rs201885728	37451744	657	T/C	L/S	Damaging(0.01)	Neutral(-0.08)	possibly damaging(0.932)	hydrophobe>hydrophile	
rs183760470	37451752	660	A/C	K/Q	Damaging(0.002)	Neutral(-0.78)	possibly damaging(0.851)	hydrophile>hydrophile	
rs41276130	37451768	665	T/G	L/W	Damaging(0.002)	Neutral(-0.98)	probably damaging(0.983)	hydrophobe>hydrophobe	
rs202200263	37454055	679	A/G	D/G	Damaging(0.001)	Neutral(-1.04)	possibly damaging(0.924)	hydrophile>hydrophobe	
rs374739457	37454063	682	G/C	E/Q	Damaging(0.018)	Neutral(-0.76)	possibly damaging(0.851)	hydrophile>hydrophile	
rs17606645	37470263	723	A/T	K/N	Damaging(0.004)	Neutral(-0.44)	possibly damaging(0.851)	hydrophile>hydrophile	
rs17590850	37470375	730	A/C	N/H	Damaging(0.002)	Neutral(-0.35)	possibly damaging(0.94)	hydrophile>hydrophile	
rs200331751	37478422	817	G/A	D/N	Damaging(0.028)	Neutral(-0.12)	possibly damaging(0.818)	hydrophile>hydrophile	
rs201628233	37478440	823	G/T	A/S	Damaging(0.022)	Neutral(-0.44)	possibly damaging(0.841)	hydrophobe>hydrophile	
rs184702413	37481992	838	A/G	E/G	Damaging(0.017)	Neutral(-1.16)	possibly damaging(0.851)	hydrophile>hydrophobe	
rs200114350	37486388	899	A/G	N/S	Damaging(0.006)	Neutral(-1.64)	possibly damaging(0.713)	hydrophile>hydrophile	
rs199691521	37488715	926	A/T	E/V	Damaging(0.02)	Neutral(-1.16)	probably damaging(0.994)	hydrophile>hydrophobe	
rs376116213	37505157	973	G/A	R/K	Damaging(0.004)	Neutral(-2.23)	probably damaging(0.976)	hydrophile>hydrophile	
rs1200875	37505192	985	C/T	R/C	Damaging(0.001)	Neutral(0.32)	possibly damaging(0.762)	hydrophile>hydrophile	
rs375945698	37505306	1023	G/C	E/Q	Damaging(0.018)	Neutral(-2.17)	probably damaging(0.999)	hydrophile>hydrophile	
rs368559588	37508121	1161	G/A	A/T	Damaging(0.04)	Neutral(-0.75)	possibly damaging(0.618)	hydrophobe>hydrophile	
rs199795040	37508139	1167	C/A	Q/K	Damaging(0.002)	Neutral(-2.3)	possibly damaging(0.886)	hydrophile>hydrophile	

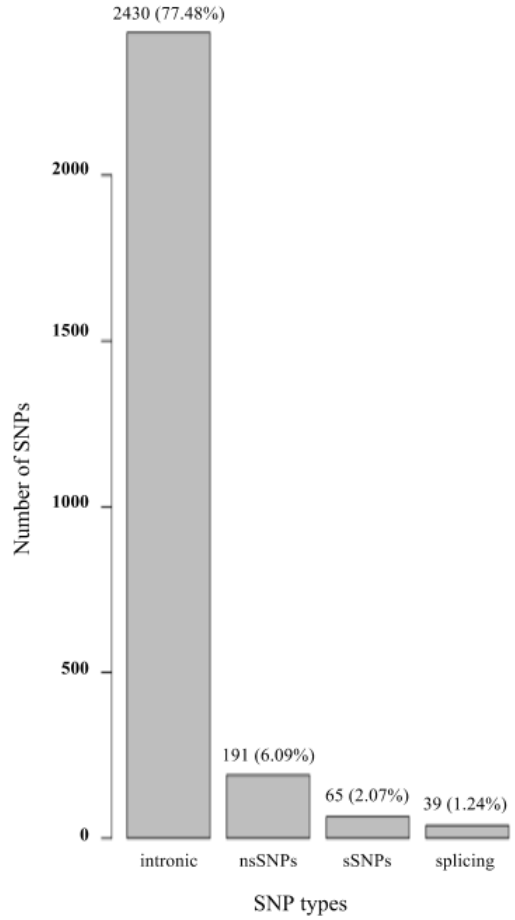


Figure 1 Graphical representation of distribution of intronic SNPs, non-synoymous SNPs (nsSNPs), synonymous SNPs (sSNPs), and SNPs at splicing sites for the NY-BR-1 gene, based on the dbSNP and NHLBI ESP databases.

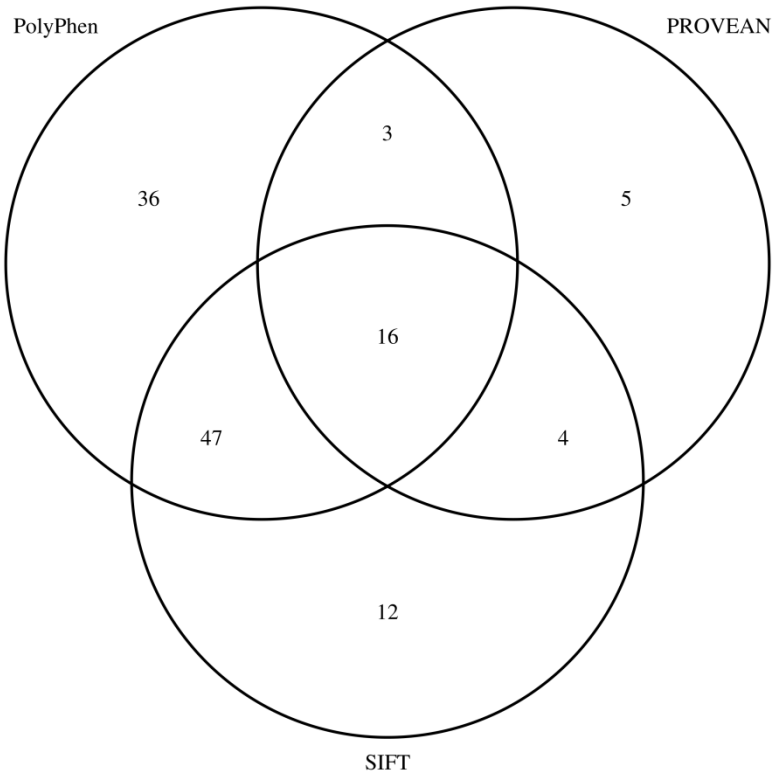


Figure 2 Venn diagram showing the overlap of the predictions made by the three tools PolyPhen, PROVEAN and SIFT.

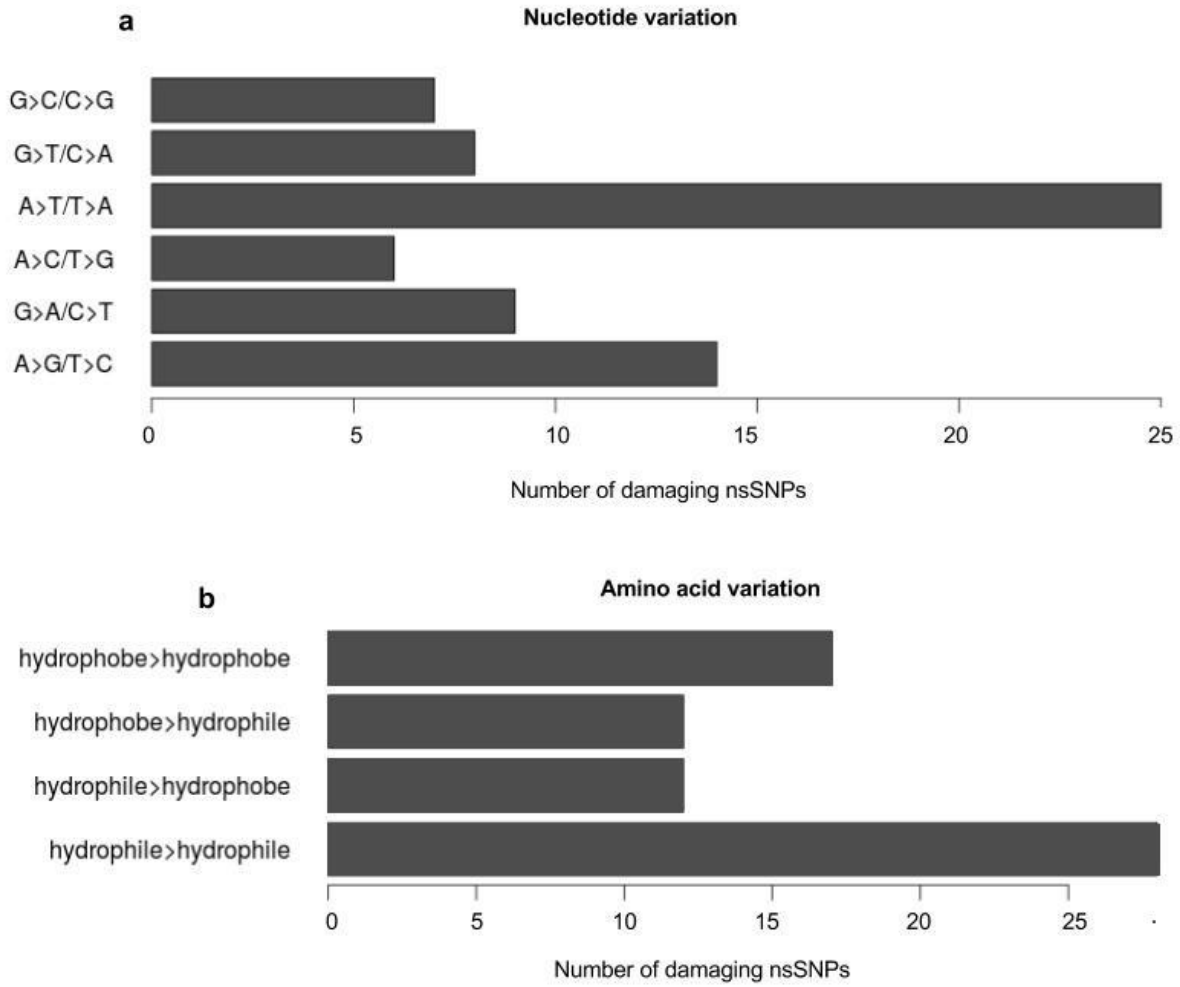


Figure 3 Graphical representation of spectrum of damaging nsSNPs variation. a) nucleotide variations, b) amino acid variations.

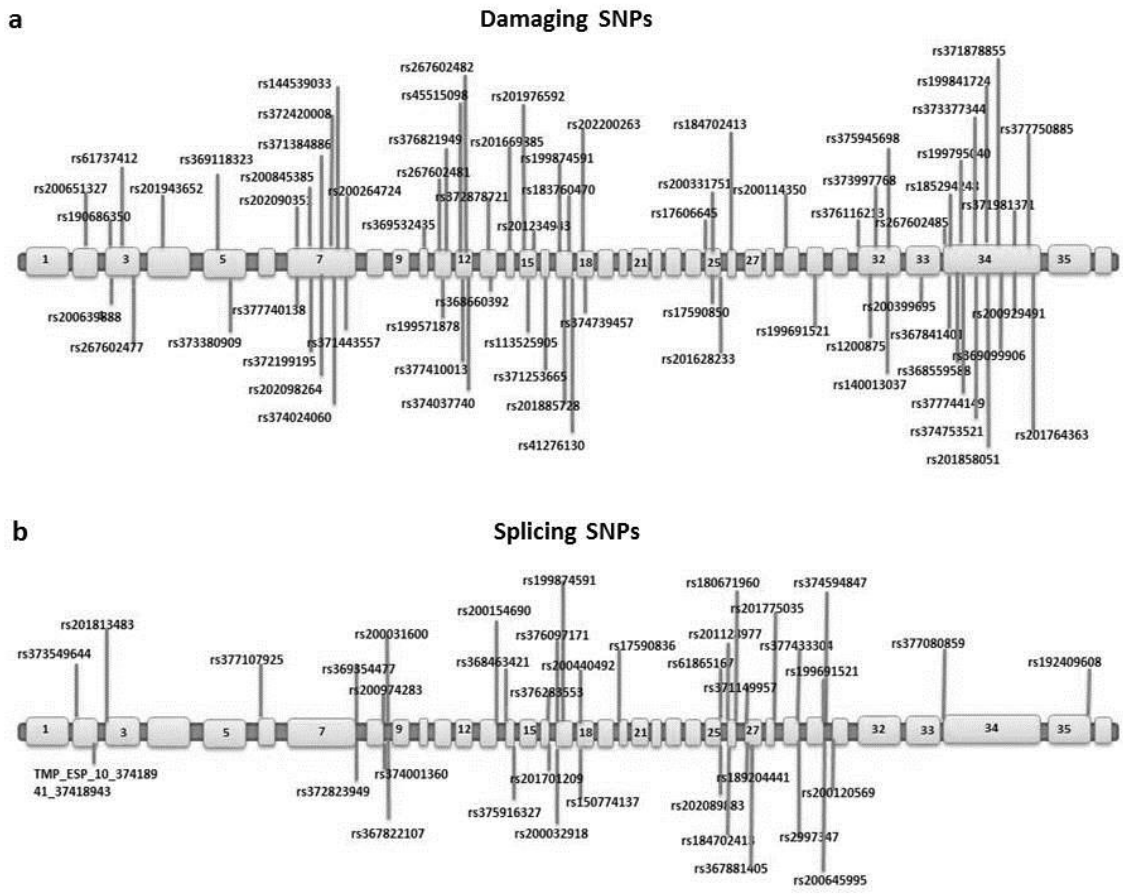


Figure 4 Graphical representation of location of NY-BR-1 SNPs. a) damaging SNPs, b) splicing SNP.