Delineating cellular heterogeneity and organization of breast cancer stem cells

Nina Akrap

Department of Pathology

Institute of Biomedicine

Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2015

Cover illustration: Micrograph of PKH-stained MCF7 mammospheres by Nina Akrap

Für Baka

Delineating cellular heterogeneity and organization of breast cancer stem cells

Nina Akrap

Department of Pathology, Institute of Biomedicine
Sahlgrenska Academy at University of Gothenburg
Göteborg, Sweden

ABSTRACT

Breast cancer is characterized by a high degree of heterogeneity in terms of histological, molecular and clinical features, affecting disease progression and treatment response. The cancer stem cell (CSC) model suggests, that cancers are organized in a hierarchical fashion and driven by small subsets of CSCs, endowed with the capacity for self-renewal, differentiation, tumorigenicity, invasiveness and therapeutic resistance. The overall aim of this thesis was to characterize CSC phenotypes and the cellular organization in estrogen receptor α + (ERα+) and ERα- subtypes of breast cancer at the individual cell level. Furthermore, we aimed to identify novel functional CSC markers in a subtype-independent manner, allowing for better identification and targeting of breast-specific CSCs.

At present, single-cell quantitative reverse transcription polymerase chain reaction represents the most commonly applied method to study transcript levels in individual cells. Inherent to most single-cell techniques is the difficulty to analyze minute amounts of starting material, which most often requires a preamplification step to multiply transcript copy numbers in a quantitative manner. In **Paper I** we have evaluated effects of variations of relevant parameters on targeted cDNA preamplification for single-cell applications, improving reaction sensitivity and specificity, pivotal prerequisites for accurate and reproducible transcript quantification.

In **Paper II** we have applied single-cell gene expression profiling in combination with three functional strategies for CSC enrichment and identified distinct CSC/progenitor clusters in ERα+ breast cancer. ERα+ tumors display a hierarchical organization as well as different modes of cell transitions. In contrast, ERα- breast cancer show less prominent clustering but share a quiescent CSC pool with ERα+ cancer. This study underlines the importance of taking CSC heterogeneity into account for successful treatment design.

In **Paper III** we have used a non-biased genome-wide screening approach to identify transcriptional networks specific to CSCs in ERα+ and ERα- subtypes. CSC-enriched models revealed a hyperactivation of the mevalonate metabolic pathway. When detailing the mevalonate pathway, we identified the mevalonate precursor enzyme 3-hydroxy-3-methylglutaryl-CoA synthase 1 (HMGCS1) as a specific marker of CSC-enrichment in ERα+ and ERα- subtypes, highlighting HMGCS1 as a potential gatekeeper for dysregulated mevalonate metabolism important for CSC-features. Pharmacological inhibition of HMGCS1 could therefore be a novel treatment approach for breast cancer patients targeting CSCs.

Keywords: Breast cancer, cancer stem cells, cellular heterogeneity

SAMMANFATTNING PÅ SVENSKA

Bröstcancer är den vanligaste cancerformen hos kvinnor och utgör 30% (2011) av alla cancerfall hos kvinnor i Sverige. Sjukdomen kännetecknas av stor variation och bröstcancer kan beskrivas som ett samlingsbegrepp för olika typer av cancer. Olika varianter av bröstcancer har olika sjukdomsförlopp och det finns undergrupper med bra respektive dålig prognos som behandlas på olika sätt.

En tumör består av många olika typer av celler. Flera modeller har försökt förklara anledningen till denna cellvariation varav en är cancerstamcellsmodellen. Här tror man att en liten del av cellerna i tumören, kallade cancerstamceller, är aggressiva, kan bilda metastaser och är motståndskraftiga mot behandling. Därför tror man att det är viktigt att hitta behandling riktad mot dessa celler. Syftet med detta arbete är att studera cancerstamceller i olika typer av bröstcancer och vidare titta på organisationen av dessa celler och andra celltyper i tumörerna. Ett annat mål med avhandlingen är att identifiera markörer som är specifika för cancerstamceller jämfört med andra cellpopulationer i olika typer av bröstcancer för att kunna använda dessa till att utveckla metoder för diagnos och behandling.

Cancerstamceller utgör en väldigt liten del av tumörcellerna och för att studera dessa krävs specifika metoder där man sorterar ut och analyserar enskilda celler. Enskilda celler innehåller väldigt lite material och därför måste materialet först amplifieras för att kunna analyseras med tillgängliga metoder. I artikel I har vi utvecklat en metod för att amplifiera denna typ av material på ett sätt som ger tillförlitliga resultat.

I artikel II har vi studerat organisationen av olika cellpopulationer i två typer av bröstcancer. Olika metoder användes för att anrika cancerstamceller som sedan jämfördes med vanliga cancerceller. Vi tittade på två olika typer av bröstcancer och i båda fallen identifierades en grupp av liknande cancerstamceller. I en av cancertyperna identifierades ytterligare populationer av cancerstamceller och man kunde se en tydlig organisation av olika celltyper. Denna studie påvisar betydelsen av att behandla all typer av relevanta cellpopulationer för att eliminera cancer.

I artikel III försökte vi identifiera markörer specifika för cancerstamceller i olika typer av bröstcancer. Vi använde en speciell metod för att hitta signalvägar specifika för cancerstamcellerna. Genom att titta närmare på

en av signalvägarna som upptäcktes med denna metod identifierade vi en markör kallad HMGCS1 som är viktig för funktionen av cancerstamceller. Farmakologisk hämning av HMGCS1 skulle därför kunna vara ett nytt behandlingssätt för bröstcancerpatienter.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Andersson, D*., **Akrap, N*.**, Svec, D., Godfrey, T.E., Kubista, M., Göran Landberg, G. and Ståhlberg, A. Properties of targeted preamplification in DNA and cDNA quantification Expert Rev Mol Diagn. 2015 Aug;15(8):1085-100. *Authors contributed equally.

II. **Akrap, N.**, Andersson, D., Gregersson, P., Bom, E., Anders Ståhlberg, A. and Landberg, G. Identification of distinct breast cancer stem cell subtypes based on single cell PCR analyses of functionally enriched stem and progenitor pools. Manuscript.

III. Walsh, C.A., **Akrap, N.,** Magnusson, Y., Harrison, H., Andersson, D., Rafnsdottir, S., Choudhry, H., Buffa, F.M., Ragoussis, J., Ståhlberg, A., Harris, A. and Landberg G. The mevalonate precursor enzyme HMGCS1 is a novel marker and key mediator of cancer stem cell enrichment in luminal and basal models of breast cancer. Manuscript.

CONTENT

ABBREVIATIONS

AI	Aromatase inhibitors
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
Acetyl-CoA	Acetyl-Coenzyme A
ALDH	Aldehyde dehydrogenase
ALDH1A3	Aldehyde dehydrogenase 1 family, member A3
ATP	Adenosine 5´-triphosphate
BCSC	Breast cancer stem cell
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
℃	Degree Celsius
CCNA2	Cyclin A2
CD	Cluster of differentiation
CD49f	Also known as integrin alpha-6
CDH1	Cadherin 1, type 1
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	complementary DNA
CFSE	Carboxyfluorescein succinimidyl ester
Cq	Quantification cycle
CSC	Cancer stem cell
DFS	Disease-free survival
DHCR24	24-dehydrocholesterol reductase
DLL1	Delta-like 1 (Drosophila)
DMAPP	Dimethylallyl pyrophosphate
DNA	Deoxyribonucleic acid
DNER	Delta/notch-like EGF repeat containing
e.g.	Exempli gratia
EMT	Epithelial-to-mesenchymal transition
EPCAM	Epithelial cell adhesion molecule
ER	Estrogen receptor
ERBB2	Erb-b2 receptor tyrosine kinase 2, encodes for HER2
FACS	Fluorescence-activated cell sorting
FDA	Food and Drug Administration
FDG-PET	Fluorodeoxyglucose positron emission tomography
FFP	Farnesyl pyrophosphate
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
FOXA1	forkhead box A1

G_0/G_1	Gap_0/Gap_1 cell cycle phase
GADD45	Growth arrest and DNA-damage-inducible, alpha
GATA3	GATA binding protein 3
GGPP	Geranylgeranyl pyrophosphate
GI/GII/GIII	Histological grade I-III
GRB7	Growth factor receptor-bound protein 7
GTPases	Ras and Rho small guanosine triphosphatases
HER2	Human epidermal growth factor receptor 2
HIF	Hypoxia-inducible transcription factor
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1
HRE	Hypoxic-response element
i.e.	*id est*
ID1	Inhibitor of DNA binding 1
IDC-NOC	Invasive ductal carcinoma not otherwise specified
IHC	Immunohistochemistry
IPP	Isopentylpyrophosphate
Ki67	Marker of proliferation
MAP3K1	Mitogen-activated protein kinase kinase kinase 1
MaSC	Mammary stem cell
min	Minute
MMTV	Mouse mammary virus tumor
MVA	Mevalonate
MVK	Mevalonate kinase
MYC	v-myc avian myelocytomatosis viral oncogene homolog
n	Sample size
N-BP	Nitrogen-containing bisphosphate
NANOG	Nanog homeobox
nM	Nanomolar
NOD/SCID mouse	Nonobese diabetic/severe combined immunodeficiency mouse
NPI	Nottingham Prognostic Index
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGR	Progesterone receptor
PI3KCA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
RAS	Rat sarcoma viral oncogene homolog

RB1	Retinoblastoma 1
RNA	Ribonucleic acid
RT-qPCR	Reverse transcription– quantitative polymerase chain reaction
SD	Standard deviation
SEM	Standard error of the mean
SERM	Selective estrogen receptor modulator
siRNA	Small interfering RNA
SNAI1	Snail family zinc finger 1
SOM	Self organizing map
SOX2	SRY (sex determining region Y)-box 2
TDLU	Terminal ductal lobular unit
TCA	Tricarboxylic acid
TNBC	Triple negative breast cancer
TP53	Tumor protein p53
Wnt1	Wingless-type MMTV integration site family, member 1
ZIC3	Zic family member 3

1 INTRODUCTION

1.1 The normal breast and breast cancer

1.1.1 The normal breast

Breast development

Mammary gland morphogenesis is initiated in the embryo at around four weeks. Most of the breast growth takes place at puberty under the influence of growth hormones and estrogen, leading to an enlargement of the rudimentary mammary epithelium. During pregnancy alveolar morphogenesis is induced by several hormones and the mammary epithelium undergoes rapid proliferation, resulting in increased ductal branching and the development of the alveolar epithelium, capable of milk secretion [1].

Breast structure

The mammary epithelium is characterized by a high degree of plasticity throughout life. The mature epithelium is organized into a series of branching ducts, which are lined by a bi-layered epithelium, consisting of luminal and myoepithelial/basal cells adjacent to a basement membrane. Mammary ducts are surrounded by stromal cells, such as adipocytes and fibroblasts and infiltrated with blood and lymph vessels. Each duct ends into the terminal ductal lobular unit (TDLU), which consists of ductules and alveolar buds. The majority of breast cancers arise in the TDLUs [1, 2] (**Fig.1A and 1B**).

Cellular hierarchy

Today, it is widely accepted that the mammary epithelium is organized in a differentiation hierarchy. Bipotent mammary stem cells (MaSCs) form the apex of the hierarchy, giving rise to unipotent luminal and basal stem

or progenitor cells, which maintain the terminally differentiated cell types. However, the exact definition of MaSCs and derived progenitor populations still remains a matter of debate. To interrogate the hierarchical organization of the mammary epithelium the field has broadly relied on *in vivo* and *in vitro* assays to test self-renewal and differentiation capacity in subsets of epithelial cells. MaSC of the adult gland are notoriously difficult to study due to their low frequency and the lack of appropriate markers. Data derived from these studies have been conflicting, which is likely the result of different applied tumor dissociation protocols and assays to assess 'stemness' [3]. Several studies indicated that the MaSC (*i.e.* cells with highest repopulating capacity) have an EPCAM^{low}/CD49f^{high} phenotype and are part of the basal cell compartment [4, 5], whereas other studies showed that the luminal and basal compartment contains MaSC and bi-potent progenitors [6]. Additionally, suprabasal luminal cells of the ducts were suggested to contain MaSC [7, 8]. Besides, there is also evidence for the existence of unipotent stem/progenitor cells that maintain the luminal and basal compartment. Luminal progenitor cells can be identified by their EPCAM^{high}/CD49f^{high} immunophenotype [9, 10]. No specific marker profile has yet been identified for basal progenitor cells, but they can be identified from serial passaging of MaSCs, indicating that these cells lie downstream in the hierarchy [10]. One feature of adult stem cells is their slow division cycle, which enables enrichment of these cells by label-retention methods, such as synthetic DNA nucleosides or membrane dyes. Pece and colleagues have used the lipophilic PKH26 dye in combination with the mammosphere assay to enrich for MaSC based on their quiescent nature [8]. The authors identified cells expressing the cell surface marker profile CD49f^{high}/DLL1^{high}/DNER^{high} to have the highest mammosphere-initiating potential. Interestingly, the gene signature derived from PKH26^{high} cells was able to predict biological and molecular features of breast cancers.
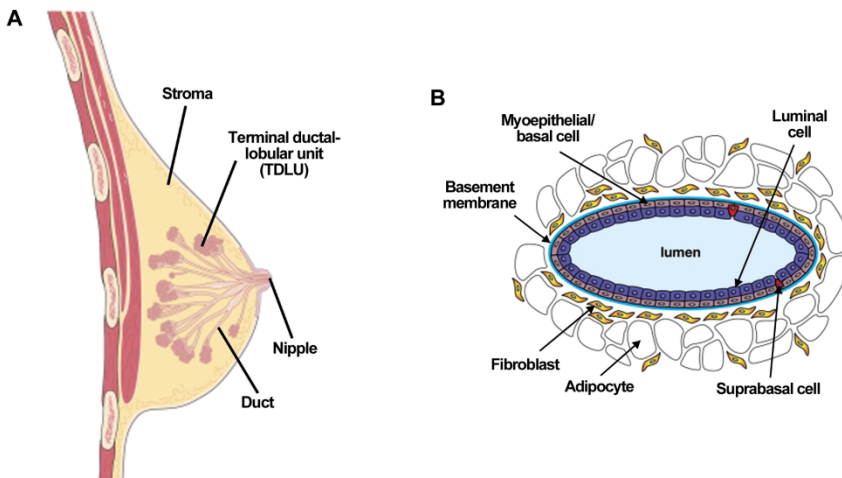
*Figure 1. Schematic illustration of the normal breast. **A:** Representation of the human mammary gland. **B:** Cross section of a mammary duct. Adapted from [2].*

1.1.2 Breast cancer

Breast cancer is the most common type of cancer diagnosed in women worldwide, with an incidence of about 25% [11] corresponding to 1.7 million women being diagnosed with breast cancer in 2012. There was a sharp rise (20%) in breast cancer incidence since 2008, which can be partly explained by changes in lifestyle common to industrialized nations [12]. Despite of the high incidence, breast-cancer related mortality is decreasing, with 5-year and 10-year survival rates of 87.8% and 78.8% in Sweden [13].

The risk of developing breast cancer has been linked to numerous factors. A few well-established risk factors comprise age, lifestyle and environmental factors, such as body mass index, alcohol consumption and hormone replacement therapy. Early menarche, late menopause and late age of first childbirth comprise additional risk factors. Most of the breast cancers are sporadic and non-familial. Hereditary forms of cancer only constitute 5-10% of all cancers. However, female carriers of germline mutations in high penetrance genes, such as *BRCA1* and *BRCA2* present a 60-80% lifetime risk of developing breast cancer [14]. Additional high penetrance mutations in for example the *PTEN* gene (Cowden syndrome) or *TP53* (Li-Fraumeni syndrome) are associated with a significant

increased risk of breast cancer. Mutations in these susceptibility alleles are rare in the general population and only account for a small fraction of susceptibility for breast cancer [14].

1.1.3 Breast cancer subtypes

Breast cancer has long been perceived as a complex disease, reflected in diverse morphological, clinical and molecular characteristics. Traditionally breast cancer is classified according to histopathological features, involving tumor size, nodal status and metastasis, also referred to as the TNM staging system. In addition, immunohistochemical parameters, such as estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) status as well as proliferation-associated markers (*e.g.* Ki67) are routinely assessed to classify breast cancers and to guide appropriate treatment decisions. More recently, with the invention of microarray-assisted gene expression profiling, breast cancers have been grouped into distinct molecular subgroups.

Histological classification

Histological grade and histological type are two clinical parameters used to classify breast cancers into subgroups. Histological grade assesses the degree of differentiation, whereas the histological type signifies the growth pattern of the tumor. The most common type of breast carcinoma are invasive ductal carcinoma not otherwise specified (IDC-NOC), which accounts for 50-80% of all carcinomas, followed by invasive lobular carcinomas accounting for about 5-15% of all cases. The remaining cases of invasive carcinomas comprise at least 17 histological types [15].

TNM staging and the Nottingham Prognostic Index

In breast cancer a useful prognostic factor ideally separates groups of patients who require no further adjuvant therapy after local surgery from those patients with poor prognosis for whom additional therapy may potentially be beneficial. No single prognostic factor meeting these criteria has been identified [16]. To predict patient outcome and assist

clinical decision making several methods have been developed, such as the St. Gallen consensus criteria, the National Comprehensive Cancer Network guidelines, Adjuvant! Online and the Nottingham Prognostic Index (NPI). The latter is widely used in clinical practice to stratify the prognosis of patients. The NPI comprises three prognostic factors, the presence of lymph node metastasis, tumor size and histological grade, assembled in a prognostic index formula [17]. Numerical NPI values can be used to stratify patients into good, moderate and poor prognostic groups. However, it has been noted that the NPI does not expose the complete clinical heterogeneity and thus would benefit from taking additional parameters into account to improve personalized management of breast cancer patients [18].

Immunohistochemical classification

In addition to the above-described histopathological parameters, ER, PR and HER2 are used as prognostic, but mainly as predictive markers, guiding treatment strategies. ER and PR status have been used for many years to assess if patients are suitable for endocrine therapy. ER is a transcription factor and required for estrogen-stimulated growth. About two thirds of all breast cancers express ER. PR expression is regulated by estrogens and therefore its expression is thought to indicate a functioning ER pathway, which may assist in predicting response to endocrine therapy. Immunohistochemistry (IHC) is the standard method for determination of hormone receptor status. Levels of ER and PR immunreactivity can be assessed using the Allred scoring system, combining scores for intensity and the proportion of cells stained. Patients may be suitable for endocrine therapy with only 1-10% of positively stained nuclei. The oncogene *ERBB2* encodes for HER2, a member of the epidermal growth factor family of tyrosine kinases. *ERBB2* is located on chromosome 17q21 and its gene product is involved in cell differentiation, adhesion and motility. The predominant mechanism of overexpression in breast cancer is gene amplification, occurring in about 20% of all breast cancers. HER2 expression is used as a predictive marker for specific systematic therapy with the humanized monoclonal antibody trastuzumab. HER2 assessment is conducted by IHC and *in situ* hybridization. The IHC score takes staining intensity and the percentage of positive cells into account. Patients presenting more than 10% of highly stained cells are qualified for targeted treatment. Borderline samples undergo further

assessment by *in situ* hybridization, applying a dual probe set that targets the centromere of chromosome 17 as well as the *ERBB2* gene locus. Individuals exhibiting an *ERBB2* to chromosome 17 ratio larger than two are suitable for HER2-specific therapy [19, 20].

Molecular classification

Microarray-based gene expression profiling studies have allowed detailed insights into the significant degree of heterogeneity of breast cancer [21-23]. These studies led to the concept that breast cancer comprises multiple diseases, affecting the same organ side and originating from the same anatomical structure (*i.e.* the TDLU), but display differences in risk factors, clinical behavior, histopathological features and response to therapy [24]. By using hierarchical cluster analysis the seminal studies by Perou *et al.,* (2000) and Sorlie *et al.,* (2001) have revealed the presence of at least four molecular groups. In addition, these studies demonstrated that ER+ and ER- breast cancers are distinct diseases at the molecular level and that the observed clusters were mainly contributed to differential expression of ER and ER-related genes, proliferation-associated genes and to a lesser extend to HER2 and genes mapping the region of the *HER2* amplicon. Today, at least six different molecular subtypes are recognized; luminal A and B, basal-like, HER2-enriched, normal breast-like as well as the more recently discovered claudin-low subgroup [25] (**Fig.2A**). Importantly, identified subtypes are associated with differences in clinical outcome (**Fig.2B**). Specific features of individual subtypes are summarized in **Table 1**.
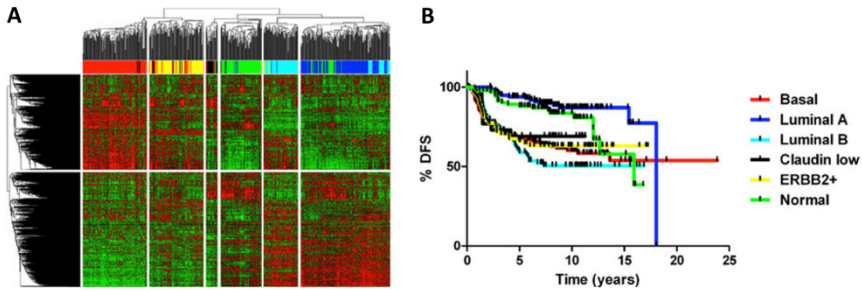
Figure 2. Human breast tumors cluster into six molecular groups and exhibit differences in survival. A: *Hierarchical clustering of 547 breast tumors into six intrinsic subtypes.* ***B:*** *Kaplan-Meyer survival analysis of the six distinct breast cancer subtypes. DFS, disease-free survival. Adapted from [26].*

Table 1. Features of microarray-based defined molecular subtypes of breast cancers. *Adapted from [24].*

	IHC markers	Histological Grade	Proliferation	Other markers	Outcome
Luminal A	ER+: 91-100 % PR+: 4-53% HER2+: 15-24% Ki67: high Basal markers: -	GI/GII: 70-87% GIII: 13-30%	Low	FOXA1 high	Good
Luminal B	ER+: 91-100 % PR+: 70-74% HER2+: 8-11% Ki67: low Basal markers: -	GI/GII: 38-59% GIII: 41-62%	High	FGFR1 and ZIC3 amp	Intermediate or poor
Basal-like	ER+: 0-19% PR+: 6-13% HER2+: 9-13% Ki67: high Basal markers: +	GI/GII: 7-12% GIII: 88-93%	High	RB1: low/- CDKN2A: high BRCA1: low/- FGFR2: amp	Poor
HER2-enriched	ER+: 29-59% PR+: 25-30% HER2+: 66-71% Ki67: high Basal markers: -/+	GI/GII: 1-45% GIII: 55-89%	High	GRB7: high	Poor
Normal Breast-like*	ER+: 44-100% PR+: 22-63% HER2+: 0-13% Ki67: low/intermediate Basal markers: -/+	GI/GII: 37-80% GIII: 20-63%	Low/intermediate	-	Intermediate
Claudin-low	ER+: 12-33% PR+: 22-23% HER2+: 6-22% Ki67: intermediate Basal markers: -/+	GI/GII: 23-62% GIII: 38-77%	Intermediate/high	CDH1: low/- Claudin: low/-	Intermediate

*At the RNA level, breast cancers of this subtype show noticeable similarities with normal breast tissue and fibroadenomas. It has recently been suggested that this subtype represents an artifact due to sample contamination with stromal, inflammatory and normal breast cells [24].

Although the different molecular subtypes are now well recognized, there are still limitations in regards to the definition and number of subtypes, and their prognostic and predictive significance. Furthermore, the information received from gene expression profiling beyond ER, PR, HER2 and proliferation markers remains to be fully established [24].

1.1.4 Breast cancer therapy

The majority of breast cancers in the developed parts of the world are diagnosed at early stage of the disease, owing to population-wide mammogram screenings. Early stage breast cancers can be completely resected by surgery followed by adjuvant therapy to prevent recurrence, which has been the gold standard in breast cancer for a long time. More recently, neoadjuvant treatment has been introduced and is clinically indicated for patients with large tumor size and high nodal involvement or patients presenting an inflammatory component [27].

Therapy for hormone receptor positive breast cancers

Hormone receptor (estrogen and progesterone) positive breast cancers constitute up to 65-75% of all breast cancers [28]. For growth and survival, hormone receptor positive breast cancers largely depend on hormone supply, which is essential for endocrine treatment design. Although, hormone receptor positive breast cancers are associated with the best prognosis amongst all subtypes, 20% of the patients experience recurrence within 10 years after surgery. The two main adjuvant modalities currently provided are cytotoxic chemotherapy and endocrine therapy, both leading to an improvement of disease-free and overall survival. There are two main classes of endocrine therapy agents; selective estrogen receptor modulators (SERMs) and aromatase inhibitors (AIs). SERMs bind to estrogen receptors in a competitive fashion to inhibit DNA synthesis by recruitment of co-repressors and inhibition of G_0/G_1 cell

cycle progression. The most commonly applied drugs of this class are tamoxifen, raloxifen and toremifene. AIs inhibit the enzyme aromatase, which converts circulating androgens into estrogens by an aromatization reaction, resulting in reductions of serum, tissue and tumor cell estrogen levels. AIs can exert their function only if the primary source of estrogen is eliminated, such as in postmenopausal women, after oophorectomy or in combination with estrogen deprivation therapy [27].

Therapy for HER2 positive breast cancer

HER2 overexpression is one of the most important carcinogenic features and HER2 amplified breast tumors have the second-poorest prognosis amongst breast cancer subtypes paralleled by lower disease-free and overall survival rates. About 20-25% of all breast cancer cases are characterized by overexpression of the HER2 protein, which is a prognostic and predictive marker for HER2 targeted therapy. HER2 is a transmembrane protein with an extracellular ligand-binding domain and an intracellular tyrosine kinase domain. The receptor is activated upon ligand binding, leading to homo- or heterodimerization with other HER protein family members. HER2 signaling is crucial, since it triggers the downstream activation of multiple pathways involved in cell proliferation and inhibition of apoptosis. Trastuzumab is a recombinant humanized monoclonal antibody and was the first FDA-approved targeted treatment for breast cancer, targeting the extracellular domain of HER2. Clinical studies have highlighted that combined treatment of trastuzumab with standard chemotherapy produces improved response rates compared to chemotherapy alone [29].

Therapy for triple negative breast cancer

Triple negative breast cancer (TNBC) is characterized by the lack of ER, PR and HER2 expression and accounts for about 10-15% of all breast cancers, frequently occurring in younger and African women as well as in BRCA-mutated individuals. TNBC represent a highly heterogeneous group of tumors and survival of patients with metastatic or recurrent disease remains poor. Given the lack of effective drug targets,

chemotherapy is used as the standard therapy, which is however more beneficial than in hormone-receptor positive breast cancers [27].

Personalized breast cancer treatment

Personalized medicine aims to classify individuals into subgroups that differ in their response to a specific treatment. With the advance of gene-expression profiling, several multi-gene expression tests for determination of risk relapse in early stage breast cancer have become clinically available. Molecular diagnostic tests include for example MammaPrint® (Agendia), Oncotype DX® (Genomic Health) and PAM50® (Prosigna), using RT-PCR or microarray technology. MammaPrint is a microarray-based gene expression profiling test, analyzing 70 genes involved in cell cycle regulation, angiogenesis, invasion, metastasis and signal transduction. The test stratifies patients into low- or high-risk groups of distant recurrence and proved to be a robust predictor for distant metastatic-free survival, independent of adjuvant treatment, tumor size, histological grade, and age. Oncotype DX uses a 21-gene expression signature to generate a prognostic parameter, termed recurrence score, predicting the risk of distant recurrence in node-negative ER+ breast cancer patients treated with tamoxifen. Based on the obtained gene signatures, patients are classified into low, intermediate and high-risk groups. Similarly, the PAM50 test uses a 58-gene signature to stratify patients into low, intermediate and high-risk groups [30, 31].

1.2 Tumor heterogeneity

Breast cancer comprises a diverse group of neoplasms originating in the epithelial cells of the mammary ducts. Heterogeneity exists between different tumors (inter-tumor heterogeneity) as well as at the individual tumor level (intra-tumor heterogeneity) [32].

1.2.1 Inter-tumor heterogeneity

Clinical traits that differ amongst breast cancers include proliferation rate, invasiveness, metastatic potential and response to treatment [33]. Several

hypotheses have been developed to explain the underlying reasons for intertumoral heterogeneity, such as different cells of origin as well as different oncogenic events. Each breast cancer results from an accumulation of oncogenic hits in a genetically normal cell. During the early stage of tumor progression clonal expansion critically determines the behavior and progression of the resulting tumor. It is thought that characteristics of the cell of origin are epigenetically conveyed to the tumor cells and their progeny [33].

DNA and exome sequencing technologies have enabled large-scale studies of breast cancer cohorts. Comprehensive molecular analyses revealed associations between tumor subtypes and sets of mutated genes [34, 35]. An extensive and integrated study by the Cancer Genome Atlas Network [35] included 852 primary breast cancer patients, which were analyzed by genomic DNA copy number arrays, DNA methylation, exome sequencing, mRNA arrays, microRNA sequencing and reverse-phase protein arrays. The authors found breast cancers to congregate into four phenotypically different classes (luminal A, luminal B, basal and HER2 amplified) due to distinct genetic and epigenetic changes. The lowest mutational rates were identified in luminal A tumors, whereas basal and HER2 amplified tumors exhibited the highest mutational rate. Mutated genes were shown to differ between subgroups, luminal A tumors most frequently displayed mutations in *PI3KCA* (45%), *MAP3K1* (13%), *GATA3* (14%), *HER2* amplification was detected in 80% of the HER2 class along with a high frequency of *TP53* (72%) and *PIK3CA* (39%) mutations, while basal tumors were characterized by high *TP53* (80%) mutations. Interestingly, intrinsic tumor subtypes were not only denoted by different mutation frequencies, but also by different mutational types. For example alterations in *TP53* were mainly nonsense and frame-shift mutations in basal tumors, but missense mutations in luminal A tumors. This and other studies have underlined significant differences in the mutational profile of breast cancer subtypes and potential subtype-specific oncogenic drivers.

A second and equally important factor in the creation of breast cancer heterogeneity is the cell of origin of a tumor and how this cell relates to the mammary epithelial hierarchy and subtypes. To address this question two primary approaches have been widely applied; firstly transgenic or conditional mouse models and secondly genetic alterations of cells and subsequent *in vivo* evaluation of their tumorigenic potential in mice [36]. MaSCs have been theorized to play an important role in breast cancer

initiation due to their long life span, enabling the stepwise accumulation of genetic mutations over time and additionally because of their inherent properties of self-renewal and lineage differentiation. Another theory is that the target cell of the oncogenic transformation is recapitulated in the phenotype of the breast cancer subtype, *i.e.* basal-like tumors would be derived from transformed basal progenitor cells and luminal-like tumors would arise from transformed luminal progenitor cells [37]. More recently however, luminal progenitor cells have been put into the spotlight as putative breast cancer initiating cells. To explore cells of origin in human cancers Keller *et al.* [6] isolated luminal cells from breast reduction tissues and introduced several combinations of oncogenes using lentiviral transduction. The derived tumors displayed luminal-like and basal-like phenotypes in immunodeficent mice, comprising much of the heterogeneity observed in sporadic breast cancers. On the other hand isolated basal cells generated metaplastic tumors that did not resemble common forms of breast cancer.

1.2.2 Intra-tumor heterogeneity

Intratumor heterogeneity refers to the coexistence of cancer cell subpopulations, displaying differences in their genetic, phenotypic or behavioral traits within a given primary tumor as well as between a primary tumor and its metastasis. Two models have been suggested to account for intratumor heterogeneity, clonal evolution and the cancer stem cell theory. Both concepts are described in more detail below.

Cell oncogenic phenotypes are determined by two components, cell-intrinsic and cell-extrinsic factors. Cell-intrinsic factors refer to inherent properties of a cell and comprise genetic as well as epigenetic aspects. In normal cells phenotypic identities are mostly always defined by non-genetic mechanisms and genetic heterogeneity is usually very low. In cancer genetic mutations underlying tumor formation can have profound impact on cell phenotype and provide options for therapeutic intervention, such as in the case of *HER2* overamplifcation. The differentiation state of a cell is regulated by epigenetic mechanisms. During tumor progression the epigenome is modified by two major sources; driver mutations acquired during tumorigenesis and stochastic alterations during tumor progression. The extent of epigenetic changes may be side-specific or global. Cell-intrinsic factors however, should be regarded in a context-dependent manner. Tumor cell behavior is influenced by

microenvironmental cues, inhibiting or promoting tumor progression. Multiple factors of the tumor microenvironment contribute to cell diversity, including blood and lymph vessels, the extracellular matrix and diverse stromal cells, such as fibroblasts and immune cells as well as secreted growth factors [38, 39] (**Fig.3**).

Phenotypic heterogeneity can be classified into two groups, deterministic and stochastic. Deterministic heterogeneity denotes the existence of multiple phenotypic states. In normal tissues deterministic heterogeneity corresponds to distinct stages in the tissue-specific differentiation hierarchy. In cancer substantial genetic and epigenetic alterations as well as an atypical microenvironment may cause an increase in deterministic heterogeneity, including phenotypic states that do normally not occur in normal tissues. Stochastic heterogeneity on the other hand, defines transient alterations in phenotypes of cells that share the same deterministic phenotypic state. These differences stem from the stochastic nature of biochemical processes and from burst-like gene expression, leading to considerable cell-to-cell variation. Besides, stochastic processes can mediate transitions between distinct deterministic phenotypic states. According to the cancer stem cell (CSC) concept the phenotypiheterogeneity in cancers reflects the differentiation hierarchies present in normal tissues [40]. Phenotypic heterogeneity appears to be dominant over the effects of oncogenic transformations as shown by gene expression profiles of more differentiated and stem-like subpopulations in breast cancer cluster more closely to their respective counterparts in normal tissues then they do to each other. Furthermore, phenotypic heterogeneity has been associated to important clinical parameters, such as prognosis, treatment resistance as well as metastatic potential [41-43].
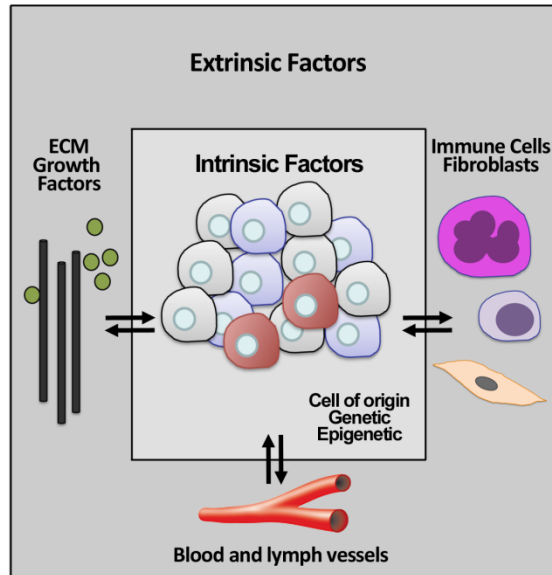
Figure 3. Determinants of tumor cell heterogeneity. *Cell-intrinsic and cell-extrinsic factors affect cellular diversity in solid tumors. Intrinsic factors comprise the biology of the cell of origin as well as genetic and epigenetic elements. Extrinsic factors arise from the microenvironment, encompassing the composition of the extracellular matrix, blood and lymph vessel supply and the recruitment of stromal cells supporting tumor growth.*

1.3 The clonal evolution theory and the cancer stem cell hypothesis

1.3.1 The clonal evolution theory

The clonal evolution theory provides a mechanism to account for intra-tumor heterogeneity and is focused on random mutations and clonal selection. According to this paradigm cancer cells in a tumor acquire several combinations of mutations over time. Eventually due to stepwise natural selection for the fittest clone, most aggressive cells drive tumor progression. The clonal evolution model suggests that tumor initiation occurs in a single cell following the acquisition of multiple mutations, providing it with a selective growth advantage. During tumor progression,

genetic instability and uncontrolled proliferation permit the accumulation of further mutations and hence new characteristics, which may provide a growth advantage over other tumor cells, *e.g.* by withstanding apoptosis. In that way new cellular variant subpopulations are generated as the tumor progresses and other subpopulations may contract, thereby producing heterogeneity (**Fig.4**). Importantly, any cancer cell in a tumor can potentially become invasive and cause metastasis or develop treatment resistance and thus lead to recurrence [38]. Mutational analysis has shown the existence of multiple subclones in diverse cancers including breast cancer [44]. Moreover, breast cancers have been demonstrated to present two classes of genetic variation, monogenomic and polygenomic tumors. Monogenomic tumors contain a single major clonal subpopulation, whereas polygenomic tumors contain multiple clonal subpopulations, accounting for tumor heterogeneity [45].

1.3.2 The cancer stem cell hypothesis

An alternative and most likely supplementary concept aiming to account for the cell diversity in tumors is the CSC hypothesis, according to which phenotypic heterogeneity in cancers is a reflection of differentiation hierarchies, existing in normal tissues. The model implies a hierarchical organization of tumor cells such that a small subpopulation of CSCs form the apex of the hierarchy and give rise to more differentiated cell types and thereby establishing the cellular diversity of the primary tumor [40, 46]. Initial evidence for the existence of CSC was shown in acute myeloid leukemia, in which a minor subset of cells could induce leukemia following transplantation into immunodeficient mice [47]. In breast cancer tumor-initiating cells were first isolated by Al-Hajj and co-workers [48] based on the expression of cell surface marker $CD44^{high}/CD24^{low}/Lineage^{negative}$ profile. As few as 100 cells exhibiting this immunophenotype were able to generate tumors in immunodeficient mice and could be serially passaged and recapitulate the heterogeneity of the primary tumor. In contrast, 10,000 cells expressing the reciprocal marker profile were unable to induce tumors in mice. In follow-up studies CSCs of breast cancers have been enriched using different combinations of markers [7, 8] (**Fig.4**).

Despite of the potential applicability of the CSC model, unequivocal characterization of cancer cell phenotypes based on their differentiation states may be impeded by the distorted identity of differentiation states.

Cancer cells acquire numerous epigenetic and genetic aberrations, possibly leading to unique mutational phenotypes, which may not exactly parallel similar states in normal cells [49]. Additionally, several studies have demonstrated that CSCs can be generated from non-CSCs by induction of the epithelial-to-mesenchymal transition (EMT) [50, 51] or convert to a CSC state spontaneously [52, 53], leading to an extension of the classical CSC model to include the phenomenon of cellular plasticity (**Fig.4**). Moreover, stemness of cancer cells can be profoundly affected by the applied functional assay.
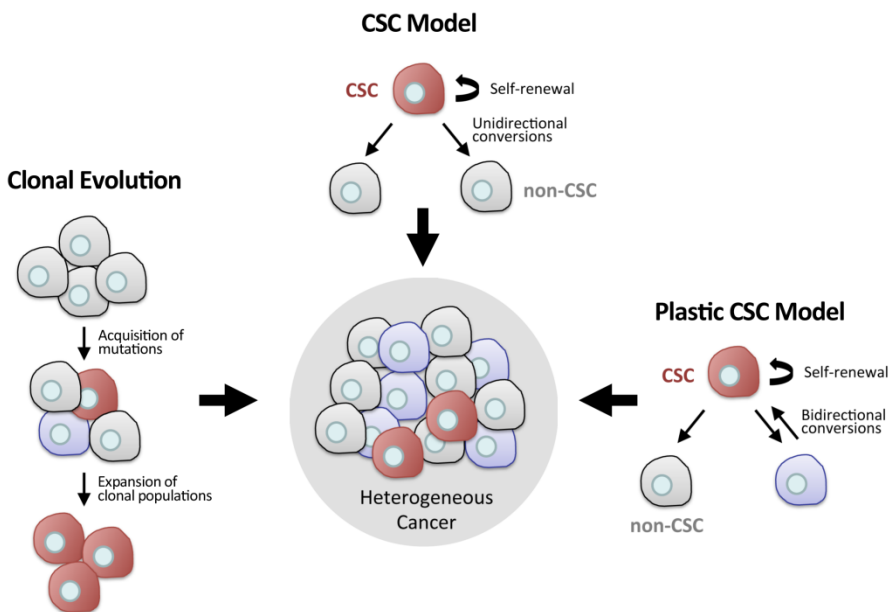

Figure 4. Clonal evolution and the CSC model create tumor heterogeneity. The clonal evolution model suggests that diverse cancer cell populations evolve during tumor progression due to the accumulation of random mutations and clonal selections, thereby contributing to tumor heterogeneity. The cancer stem cell model proposes that tumor heterogeneity arises when cancer cells reside in distinct states of stemness or differentiation within an individual tumor. In the classical CSC model conversions between cell states occur in a unidirectional manner. The plastic CSC model describes an evolving concept; according to this paradigm cell-state conversions between CSCs and non-CSCs can occur in a bidirectional fashion, implying that non-CSCs can generate CSCs throughout tumorigenesis. CSC, cancer stem cell. Modified from [39].

1.3.3 Attributes of cancer stem cells

CSC share critical features with normal tissue stem cells, including self-renewal by symmetric and asymmetric cell division and the capacity to differentiate, although in an aberrant manner. Multi-lineage differentiation however is not an obligatory feature of CSCs [46]. In addition, CSCs often use the same signaling pathways utilized by their normal counterparts, such as Notch, Wnt and Hedgehog [54]. The cancer stem cell frequency appears highly variable between different tumor types and even tumors of the same subtype. CSC numbers may change during the course of the disease and moreover CSC enumerations strongly depend on the applied assay to assess stemness, highlighting the need for more specific markers [2, 46, 55]. For the definitive identification of CSCs enriched cell fractions should re-establish the phenotypic heterogeneity of the primary tumor and exhibit self-renewing capacity on serial passaging in mouse model systems.

Besides, CSCs have been implicated in mediating metastasis [56] and increased resistance against radiation and chemotherapy, contributing to relapse following therapy [57-59]. CSC characteristics can vary across different breast cancer subtypes, for example Harrison *et al.* [60] have demonstrated that hypoxia influences CSC numbers in contrasting directions in ERα+ and ERα- breast cancer, where CSC numbers increased in the ERα+ disease following hypoxia. CSC heterogeneity has also been detected within a given tumor. Max Wicha's lab has shown that normal and malignant breast cancer stem cells express $CD44^{high}/CD24^{low}$ phenotype [48] and in addition the enzyme aldehyde dehydrogenase (ALDH) enriches for cells with CSC characteristics. In primary breast xenografts, the $CD44^{high}/CD24^{low}$ phenotype and $ALDH^{high}$ fractions identified overlapping, but non-identical cellular populations, both able to initiate tumors in NOD/SCID mice [7]. More recently the group has demonstrated that $CD44^{high}/CD24^{low}$ populations exhibit a more mesenchymal-like phenotype, whereas ALDH populations were characterized by an epithelial, proliferative phenotype [61]. Moreover, transitions between these two CSC states were found to be mediated by epigenetic mechanisms induced by the tumor microenvironment as well as transcriptional regulation. Based on their studies the authors suggested that epithelial and mesenchymal-like states of CSCs might enable these cells to invade and form distant metastasis.

1.3.4 Concluding remarks

Both, the cancer stem cell model and the clonal evolution theory are likely to exist in human cancers and are not mutually exclusive. The two concepts share certain similarities, such as the cellular origin of cancer. In both views cancer originates from an individual cell that has acquired multiple mutations and gained the potential to proliferate unlimitedly. Furthermore, consistent with both paradigms the cell of origin, genetic aberrations as well as microenvironmental factors will define the constitution of a tumor, its physical and clinical characteristics. Differences concern the mechanisms with which tumor heterogeneity is described. The CSC model proposes a program of aberrant differentiation, while the clonal evolution model suggests competition between clonal subpopulations to explain tumor heterogeneity. Furthermore, in the CSC model only a small subset of cells contribute to tumor progression, whereas any cell in a tumor has the potential to be involved in tumor progression according to clonal evolution. According to the CSC concept only CSCs may acquire further mutations which may lead to more aggressive phenotypes. Another difference concerns drug-resistance, CSC are thought to be inherently drug-resistant, while the clonal evolution models proposes a selection of drug-resistant clones [38]. These two models implicate differences in the design for new anti-cancer treatments. In the case of the CSC model, CSCs must be eradicated in order to achieve curative treatment, requiring knowledge about predominating pathways and proteins in these cell types. On the other hand, the clonal evolution model implies that effective treatment regimens should target multiple cancer cell populations.

1.4 Mevalonate pathway in cancer

1.4.1 Dysregulated metabolism in cancer

The six core hallmarks of cancer as originally postulated by Hanahan and Weinberg [62] comprise sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. Conceptual progress over the last decade has led to a revised version of the paper [63], which includes the emerging cancer hallmarks, evading immune surveillance as well as reprogramming of cellular metabolism. Cancer cells often proliferate in an uncontrolled manner with corresponding

alterations of their energy metabolism to ensure sufficient metabolite supply for cell growth and division. Normal cells under aerobic conditions metabolize glucose to pyruvate in the cytoplasm, which is then imported into mitochondria to generate adenosine 5′-triphosphate (ATP) by oxidative phosphorylation. Under anaerobic conditions pyruvate production is favored, generating ATP with a considerable lower efficiency [63].

In the 1920s Otto Warburg discovered that cancer cells, even in the presence of ample oxygen, prefer to generate ATP through glycolysis, a seeming paradox as glycolysis is less efficient in terms of ATP production compared to oxidative phosphorylation [64]. This phenomenon is called the Warburg effect, also known as aerobic glycolysis. Since then the Warburg effect has been appreciated in different types of cancers [65] and its concomitant increase of glucose uptake has been employed clinically for solid tumor detection by fluorodeoxyglucose positron emission tomography (FDG-PET). Given the low energy efficiency of the Warburg metabolism, the functional rationale so far remains unclear.

One idea to explain the Warburg effect is that glycolytic metabolism of cancer cells presents a selective advantage in the unique tumor environment. Insufficient and disorganized vessel formation in the growing tumor leads to limited blood supply, hypoxia and stabilization of hypoxia-inducible transcription factors (HIFs) [65]. HIF initiates a pleiotropic transcriptional program that counteracts hypoxic stress, including a shift towards glycolytic metabolism by upregulation of glycolytic enzymes, glucose transporters, and inhibitors of mitochondrial metabolism. With the possible exception of tumors that have lost the von Hippel-Lindau protein, HIF expression is still linked to oxygen levels, as evident from its heterogeneous expression in tumors [66, 67]. Thus, the Warburg effect cannot only be explained by HIF stabilization. Oncogene activation (*e.g. RAS*, *MYC*) and tumor suppressor loss (*e.g. TP53*, see below) have been associated with the induction of metabolic changes independently of HIFs [68]. Rapidly dividing cells require not only ATP, but also nucleotides, proteins, fatty acids and membrane lipids for biomass production. More recently, Vander Heiden *et al.*, [69] have proposed that elevated glycolysis permits the allocation of glycolytic intermediates into numerous biosynthetic pathways, enabling cells to synthesize macromolecules and organelles needed to produce a new cell. Acetyl-Coenzyme A (acetyl-CoA) for example is made available for the synthesis of several lipid building blocks, including mevalonate (MVA).

1.4.2 The mevalonate pathway for steroid biosynthesis and protein prenylation

The mevalonate pathway was discovered in the 1950s by Goldstein and Brown [70] and provides isoprenoid building blocks for the biosynthesis of diverse classes of vital cellular products, including cholesterol and prenyl pyrophosphates. The latter function as substrates for posttranslational prenylation of proteins. Imbalances of mevalonate metabolism are a well-known cause for cardiovascular diseases [71]. More recently, dysregulation of the mevalonate pathway has been implicated in various aspects of tumor development and progression [72, 73] and has been linked to CSC survival in breast cancer [74, 75].

Rapidly dividing tumor cells have high energetic requirements, in order to meet these glucose is converted into pyruvate by aerobic glycolysis as described above. Pyruvate enters the mitochondria, where it is further metabolized in the tricarboxylic acid (TCA, citrate or Krebs) cycle. However, mitochondrial oxidation is incomplete, leading to an increased export of acetyl-CoA into the cytosol, which is thereby made available for mevalonate metabolism [76] (**Fig.5**). In the mevalonate pathway thiolase condenses two acetyl-CoA molecules to produce acetoacetly-CoA. 3-hydroxy-3-methylglutaryl-CoA synthase 1 (HMGCS1) condenses acetoacetyl-CoA with another acetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). In the first committed step of the mevalonate pathway 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) converts HMG-CoA to mevalonic acid (mevalonate). HMGCR is regulated by several feedback mechanisms and the target of the cholesterol-lowering class of drugs, collectively referred to as statins. Mevalonate is then metabolized to isopentylpyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP), which both represent precursors for diverse classes of cellular products [71].

Products formed in the cholesterol branch of the pathway include steroids like estrogen, bile acids and vitamin D. In normal cells cholesterol is essential to maintain membrane integrity, modulating membrane fluidity and is involved in intracellular transport as well as in cell signaling [70]. In the prenylation branch of the pathway farnesyl pyrophosphate (FFP) and geranylgeranyl pyrophosphate (GGPP) are formed through sequential condensation reactions of DMAPP. Both, FFP and GGPP are used as adjuncts for C-terminal posttranslational modifications of various

proteins, which are referred to as proteinprenylation. Prenylation plays a role in membrane attachment and protein-protein interaction, which are essential requirements for biological functioning of proteins and is carried out by three enzymes, FTase, GGTase I and GGTase II. Prenylation occurs on many members of the Ras and Rho small guanosine triphosphatases (GTPases). The role of Ras proteins in cancer development and progression is well established [77].
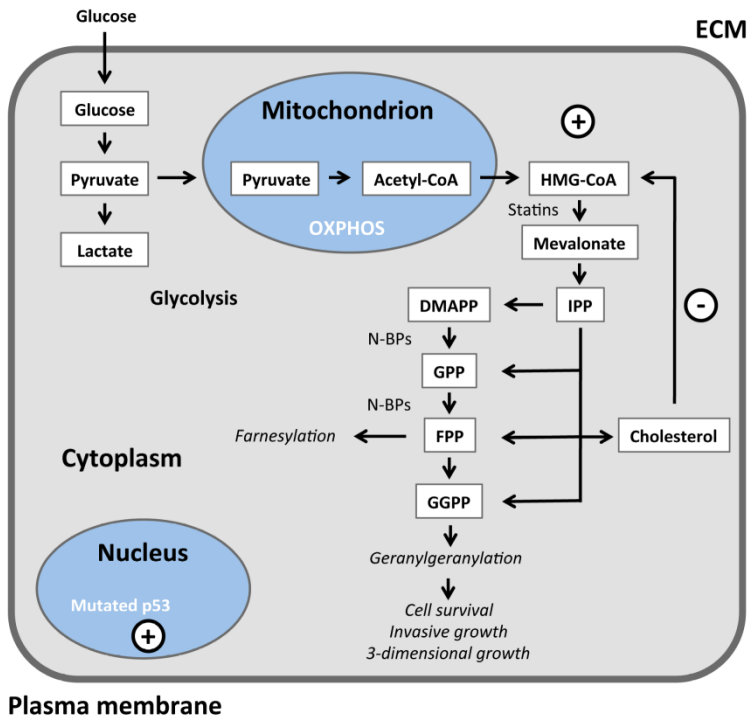

Figure 5. Metabolic reprogramming and dysregulation of the mevalonate pathway in cancer. Metabolic reprogramming of cancer cells causes upregulation of aerobic glycolysis to the expense of oxidative phosphorylation. Pyruvate is produced during aerobic glycolysis, which is either converted to lactate or further metabolized in the TCA cycle. Mitochondrial oxidation is incomplete and generates excess acetyl-CoA, which is exported into the cytosol. Cytosolic acetyl-coA can be used to generate HMG-CoA in the mevalonate metabolism, which is enhanced by certain p53 mutant variants. The mevalonate pathway can be blocked at several steps; statins inhibit HMG-CoA reductase, the first committed step of the pathway, while nitrogen-containing bisphosphates (N-BPs) inhibit FFP synthase. Modified from [76].

1.4.3 Mevalonate metabolism is regulated by mutant p53

As the "guardian of the genome" the tumor suppressor protein p53 plays an important role in the maintenance of genomic integrity and the prevention of tumor formation. p53 activation occurs through various extra- and intracellular stressors such as, DNA damage, nutrient depravation, hypoxia, oncogene deregulation, radiation or chemical agents [78]. Upon activation p53 is stabilized primarily through posttranslational modifications, which leads to its activation and accumulation in cells [79, 80]. Wild-type p53 functions as a sequence-specific, homotetrameric transcription factor, binding to degenerative DNA sequences, termed p53-responsive elements to initiate transcription of target genes. Cellular responses triggered by p53 are stimulus-dependent, *i.e.* in cells that are exposed to transient or mild stress p53 promotes cell cycle arrest (*e.g.* via p21, GADD45, 14-3-3r) or DNA repair response to facilitate cell survival. On the other hand, sustained or severe cellular stress triggers p53-mediated apoptosis (*e.g.* via Puma, Bax, Fas) and senescence (*e.g.* via p21) [81], to prevent tumorigenesis. Somatic p53 mutations are appreciated in almost all types of cancer and can be detected in more than 50% of tumors [79, 80]. In contrast to most tumor suppressors, which are inactivated following mutation, cancer-associated p53 mutations are frequently missense mutations. Single-base pair substitutions cause the translation of full-length proteins with an amino acid exchange at the respective position. These missense mutations cause p53 protein alteration and prolong the half-life of the protein. In normal, unstressed cells p53 is maintained at low levels through ubiquitination and proteasomal degradation by its negative regulators. In addition to loss of tumor-suppressive functions, certain p53 mutants can acquire novel tumor-promoting activities, referred to as gain-of-function mutations. More recent work has demonstrated that mutant p53 is also involved in many aspects of metabolic regulation in tumor cells [82, 83] Freed-Pastor *et al.* [74] have shown that mutant p53 transcriptionally enhances the expression of mevalonate pathway members by associating with sterol regulatory element-binding proteins and binding to sterol gene promoters, resulting in increased protein prenylation and maintenance of a malignant phenotype.

2 AIMS

The overall aim of this thesis is to characterize CSC phenotypes and the cellular organization in ERα+ and ERα- subtypes of breast cancer at the individual cell level. Furthermore, we have aimed to identify novel functional CSC markers in a subtype-independent manner, allowing for better identification and targeting of CSCs.

Specific aims

Paper I: Quantification of small molecule numbers frequently involves a preamplification step to generate sufficient copies for accurate downstream analyses. In **paper I** we aimed to evaluate the effects of variations of relevant parameters on targeted cDNA preamplification for single-cell reverse transcription – quantitative polymerase chain reaction (RT-qPCR) applications, to improve reaction sensitivity and specificity, pivotal prerequisites for accurate and reproducible transcript quantification.

Paper II: The large number of assays currently employed to detect CSC in breast cancer types indicates either a lack of universal markers or is reflective of the heterogenetic and dynamic nature of CSCs. In **paper II** we aimed to study the diversity of the CSC pool at the individual cell level in regards to ERα+ and ERα- subtypes, using several functional cancer stem cell enrichment techniques.

Paper III: Reliable CSC markers common to various breast cancer subtypes remain to be clearly defined and represent an essential requirement for clinical identification, monitoring and effective therapeutic targeting. In **paper III** we aimed to identify specific molecular pathways common to CSCs of ERα+ and ERα- subtypes.

3 METHODOLOGICAL ASPECTS

3.1 Single-cell qPCR

Breast cancers are complex entities, composed of heterogeneous cell types, exhibiting remarkable diversity for many tumorigenesis-related and therapy-relevant traits, such as their tumorigenic, angiogenic, invasive and metastatic potential. In addition, responses to specific treatments have been reported to differ greatly between individual tumor cells [32]. Thus, there is a vital requirement for reliable tools to scrutinize cellular behaviors at the single-cell level. One of the major limitations of conventional gene-expression profiling is that measurements are performed on composite samples, containing diverse cells in undefined proportions. Single-cell gene-expression profiling permits the identification and characterization of different cell types and furthermore, enables the correlation of gene expression patterns with phenotypical qualities, providing a comprehensive approach to assess individual cells under various conditions [84]. Inherent to most single-cell techniques is the difficulty to analyze minute amounts of starting material, which is technically more demanding. To date RT-qPCR is the most commonly applied strategy for single-cell gene-expression profiling [85]. The technique includes several sequential steps, each of which must be carefully optimized and validated. Specific steps encompass; cell collection and lysis, reverse transcription of mRNA and cDNA preamplification followed by qPCR and multivariate data analysis (**Fig.6**). Cells can be collected using various techniques, *e.g.* fluorescence-activated cell sorting or microaspiration. During the cDNA preamplification step transcript copy numbers are multiplied in a quantitative fashion, theoretically facilitating the analysis of an unlimited number of transcripts per single cell. Several preamplification approaches have been described in the literature. For single-cell applications the preferred method is targeted multiplex PCR, applying gene-specific primers. Multiplex PCR is a highly complex reaction, due to the presence of multiple primer pairs and the simultaneous amplification of large numbers of target molecules. It is critical to not introduce substantial variation or bias during the preamplification step in order to preserve the original gene expression pattern.
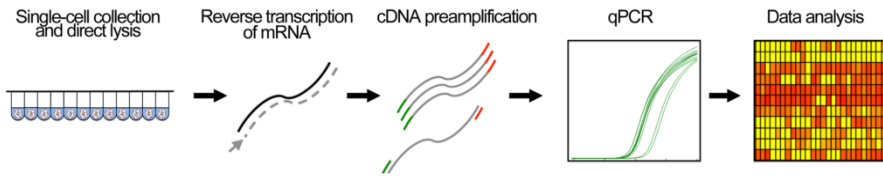
Figure 6. Workflow of single-cell qPCR. *Individual cells are collected by either fluorescence-activated cell sorting or microaspiration and lysed directly. Single-cell RNA is reverse transcribed, followed by targeted cDNA preamplification and quantitative real-time PCR. Single-cell data are typically analyzed using various uni- and multivariate statistical tools.*

3.2 Cancer stem cell enrichment methods

Investigating the role of CSCs during tumorigenesis has become a major focus in stem cell biology over the last decade. Considerable efforts have been made to develop clinical applications of the CSC model. Given the specific CSC attributes of self-renewal and differentiation, each applied marker and assay needs to be evaluated carefully [86]. The gold standard to demonstrate CSC identity is serial transplantation of cellular populations into immunocompromised mouse models. The CSC-containing population should give rise to the phenotypic heterogeneity evident in the primary tumor and demonstrate self-renewing competence upon serial passaging. The isolation of CSCs from epithelial or solid tumors is accompanied by significant technical issues, in part due to the difficulty of dissociating these tumors [2]. Furthermore, in the case of xenotransplantation incomplete immunosuppression and species-specific variations in cytokine or growth factor signaling represent confounding factors. In addition to serial passaging in mice, a number of cell surface markers have been proven useful for CSC enrichment, including CD133 (also known as prominin 1), CD44, CD24, epithelial cell adhesion molecule (EPCAM) or CD49f (also known as α6-integrin). Other CSC assays involve the Hoechst$_{33342}$ side population sorting, which is conferred by the ABC transporter ABCG2 and the ALDEFLOUR assay, based on the activity of the detoxifying enzyme aldehyde dehydrogenase, catalyzing the oxidation of retinol to retinoic acid. CSCs have frequently been enriched using markers specific for stem cells of the same organ. However, the utility of CSC markers is limited by variations in expression, regulation by environmental factors and moreover isolated

25

CSC fractions may contain considerable numbers of non-CSCs [87]. Therefore, definitive enrichment of CSCs necessitates functional assays. To circumvent obstacles associated with immunophenotypic CSC-isolation, in this work we have applied three different assays to functionally enrich for CSCs; growth in anchorage-independent culture, hypoxic culture and label-retention. Each method is explained in more detail below.

3.2.1   Growth in anchorage-independent culture

Cell culture in non-adherent conditions was originally adapted to normal breast tissue derived from reduction mammoplasties [88]. Mammary stem and progenitor cells are equipped with the unique feature of withstanding anoikis in serum-free suspension culture and generate spherical colonies, termed mammospheres. These mammospheres were found to be enriched in stem and progenitor cells. Moreover, mammosphere-derived cells differentiated along the three mammary epithelial lineages, clonally produced functional structures in 3D culture systems and reconstituted mammary glands in mouse model systems. The mammosphere assay has subsequently been adapted for quantification of stem cell activity and self-renewal capacity in cancer research and has been applied to enrich for CSC-like cells in ductal carcinoma *in situ* [89], invasive ductal carcinoma [90] and breast cancer cell lines [91]. As an example, Ponti *et al.* [90] have demonstrated that breast cancer cell-derived spheres displayed an increase in the Hoechst$_{33324}$ side population fraction, CD44^{+}/CD24^{-} cells, expressed the pluripotency-associated transcription factor OCT4 and showed high tumorigenic potential in mice. Hence, the mammosphere assay provides a functional *in vitro* tool to discover and scrutinize pathways implicated stem/progenitor cell survival.

3.2.2   Hypoxic culture

Hypoxia is commonly present in solid breast cancers and linked to malignant progression, invasion, angiogenesis, changes in metabolism and increased risk of metastasis and consequently to impaired patient prognosis. Several factors are known to cause intratumoral hypoxia, such as inadequate vascularization, an increase in diffusion distances that is associated with tumor expansion as well as tumor or therapy-related anemia. Cancer cells are able to adapt to a low-oxygen environment,

which contributes to a more malignant cellular phenotype [92]. The adaption to hypoxia is controlled by many factors, *e.g.* transcriptional and post-transcriptional changes in gene expression. In this regard, 1.5% of the human genome has been estimated to be responsive to hypoxia [93]. HIF-1α is the master regulator of the hypoxic response at the cellular level. Under hypoxia HIF 1α is stabilized and translocates to the nucleus, where it binds to the HIF 1β subunit and the co-activator p300 to activate the transcription of target genes, by binding to the hypoxic-response elements (HRE). HIF-responsive genes are involved in numerous cellular processes, including proliferation, survival, metabolism, angiogenesis, invasion and metastasis, pH regulation and the maintenance of stem cells. Moreover, cross-talks between the estrogen and hypoxic signaling pathways have been reported in breast cancer [94-96]. Under hypoxic conditions HIF-1α facilitates ERα down-regulation by proteasomal degradation as well as transcriptional repression of ERα expression [97, 98]. Several lines of evidence have reported a change of gene expression towards a more immature phenotype or an increase of cells with CSC features in response to hypoxia in different cancer types [99-101]. Furthermore, it has recently been demonstrated that hypoxia leads to increased CSC numbers in ERα+ breast cancers [60].

3.2.3 Label-retention

A less well-studied feature of CSC is cellular quiescence or dormancy, which is characterized by a low metabolic activity and entrance into a reversible G_0-G_1 arrest [102]. Various studies have used lipophilic fluorescent dyes, such as carboxyfluorescein succinimidyl ester (CFSE) or the PKH dye as well as BrdU-label retention to isolate slow-cycling breast cancer cells [8, 103, 104]. Interestingly, the work of Fillmore and Kuperwasser [103] has shown, that slow-cycling cells are present in the CD44[+]/CD24[-]/EPCAM[+] population of breast cancer cells, suggesting that these cells represent a specific CSC subset. It has furthermore been demonstrated that slow-cycling cells exhibit increased xenobiotic efflux mediated by ABCG2 transporters and increased DNA repair mechanisms [105, 106]. Taken together, these findings indicate that quiescent CSC putatively represent a small cellular subpopulation, which could be associated with resistance to chemo- and radiotherapy, disease recurrence and the formation of distant metastasis. In the **paper II** we have combined PKH26 labeling with the mammosphere assay to functionally enrich for

quiescent CSC-like cells [8]. The approach is based on the principle that during mammosphere growth, quiescent or slow dividing cells will retain the PKH26 dye, whereas the bulk population of transiently proliferating progenitor cells loses the dye due to successive cell divisions.

4 RESULTS AND DISCUSSION

4.1 Results and discussion paper I

The purpose of the preamplification is to multiply transcript copy numbers in a quantitative manner. Although several preamplification strategies have been described [107-109], for single-cell gene expression profiling, the preferred method is targeted multiplex PCR, using gene-specific primers [85]. In paper I we aimed to evaluate several experimental parameters in targeted preamplification and their effects on the reproducibility, specificity and efficiency of RT-qPCR. Specifically, variations in numbers of primers present in the multiplex reaction, primer concentrations, annealing temperature and time, cDNA template concentrations as well as the effect of PCR additives were studied (**Tab.2**). To assess its overall performance, we monitored the preamplification reaction in real-time using the DNA-binding dye SYBR Green I followed by melting curve analysis, referred to as analysis of preamplification. By using a non-specific reporter dye this method allowed us to quantitatively assess overall product formation as well as the ratios of specific and non-specific PCR products, evaluating the shape of the melting curves. Furthermore, the formation of specific amplicons was analyzed with standard RT-qPCR (**Fig.7**).

For the evaluation of targeted preamplification we optimized 96 individual PCR assays and purified and quantified the corresponding PCR products for standardization of template molecule numbers.
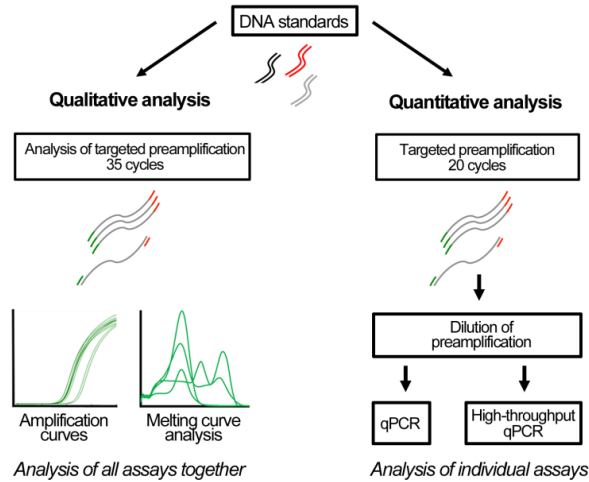
Figure 7. Experimental strategy to evaluate parameters on targeted preamplification. Left: Analysis of preamplification. *To evaluate the overall performance of targeted preamplification the reaction was monitored in real-time using SYBR Green I detection chemistry over 35 PCR cycles followed by melting curve analysis. Total product formation was quantified via amplification curves, whereas ratios of specific versus non-specific PCR product formation were derived from melting curve analyses.* **Right: Analysis of individual assays.** *Individual assays were assessed by downstream RT-qPCR following 20 cycles of preamplification, applying conventional or high-throughput RT-qPCR*

Table 2. Summary of analyzed parameters for targeted preamplification.

Tested parameter	Final primer concentration (nM)	Number of assays	Annealing time (min)	Annealing temperature (°C)	Template concentration
Template concentration (Figure 2)	40	96	3	60	6x 100 molecules + 90x 0 - 10^7 molecules
Template concentration (Figure 3)	40	96	3	60	95x 100 molecules + 1x 100 - 10^9 molecules
Number of assays (Figure 4)	40	6, 12, 24, 48, 96	0.5, 3, 8	60	96x 100 molecules
Primer concentration (Figure 5)	10, 40, 160, 240	96	0.5, 3, 8	60	96x 100 molecules
Annealing temperature (Figure 6)	40	96	3	55.0, 56.8, 58.8, 61.3, 63.3, 65.3	96x 100 molecules
PCR Additives (Figure 7)	40	96	3	60	96x 100 molecules

Theoretical molecule and preamplification cycle numbers and the dynamic range of targeted preamplification

The required number of preamplification cycles depends on the downstream qPCR platform and is primarily determined by the reaction volume, the initial cDNA concentration present in the sample as well as the dilution factor after preamplification and the preamplification efficiency [85]. In qPCR, the Poisson distribution can be applied to model the probability that a reaction chamber contains a particular number of target cDNA molecules. The variation across reaction chambers attributable to the Poisson noise leads to considerable uncertainty in the measured Cq values. Theoretically, an average of 5 molecules per reaction chamber will yield a 99.3% probability that a reaction chamber contains at least one molecule. To reduce the variation in Cq due to the Poisson effect below the variation observed for typical qPCR an average of 35 molecules is needed [85, 109]. Considering the dilution factor and the effect of Poisson noise, for 5 initial molecules, we calculate 19 cycles of preamplification to produce an average of 5 molecules per reaction chamber on the applied BioMark high-throughput qPCR, assuming a preamplification efficiency of 80%. In this study, our optimized assays displayed a preamplification efficiency of approximately 100%, which results in an average of 36 molecules per reaction chamber.

To assess the dynamic range of the preamplification we conducted two experiments, to determine the effect of total template concentrations as well as the effect of only one highly concentrated template. In the first experiment templates of 6 assays were kept at a constant concentration of 100 molecules each, whereas the remaining 90 templates were varied, ranging from 0 to 10^7 molecules per reaction. In the second experiment the initial target concentration of 95 assays was kept constant at 100 molecules per reaction and only one target was varied between 10^0 to 10^9 molecules (**Fig.8**).
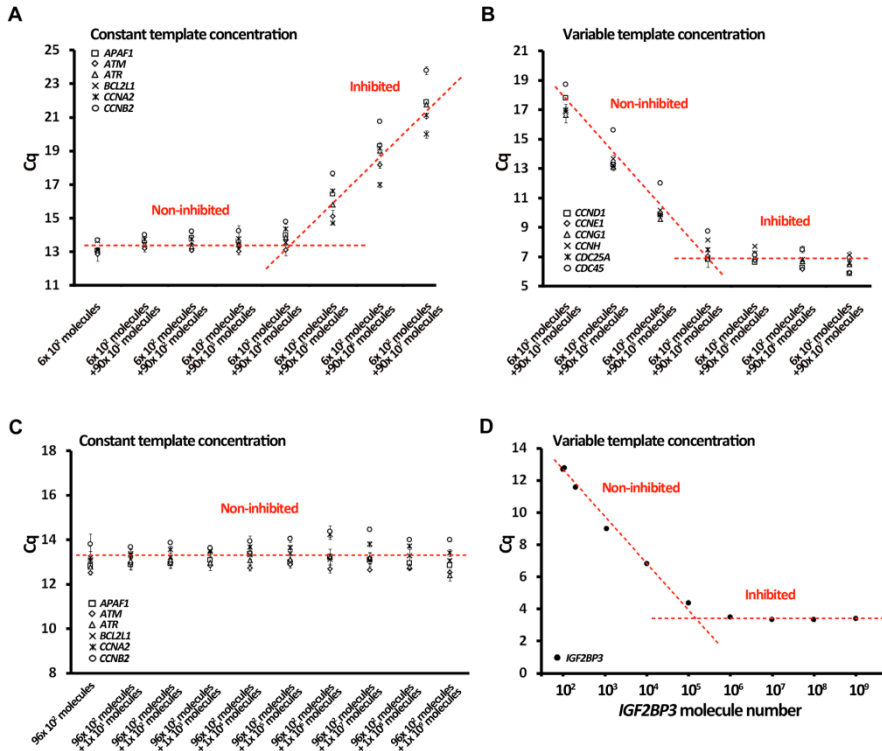
Figure 8. Dynamic range of preamplification – Effect of varied template concentrations. A. *Average Cq ±SD (n=3) of the six assays kept at a constant initial template concentration of 100 molecules each per reaction.* ***B.*** *Average Cq ±SD (n=3) of six randomly selected assays from the preamplification with an initial template concentration of 0 to 10^7 molecules each.* ***C.*** *Average Cq ±SD (n=3) of six randomly selected assays from the preamplification used at a constant initial concentration of 100 molecules each per reaction.* ***D.*** *Average Cq ±SD (n=3) of the single assay included in the preamplification with an initial template concentration of 10^2 to 10^9 molecules. The linear fit is to guide the eye only.*

For our specific reaction conditions, the preamplification was within dynamic range when the 90 templates were initially present at concentrations $<10^4$ molecules, while the remaining six templates were kept at 100 molecules per reaction (**Fig.8A and B**). Inhibition occurred at template concentrations $>10^4$ molecules. However, when only the concentration of one target molecule was increased, the remaining assays were unaffected. In summary, the preamplification dynamic range of an

assay was dependent on the amount of its target molecules and on the total number of target molecules for all the preamplification assays.

Dependence on assay numbers

To test the effect of different numbers of assays present in the preamplification, we conducted experiments containing 6, 12, 24, 48 and 96 primer pairs at a constant primer concentration of 40 nM. Analysis of preamplification showed an increase of the total PCR product yield with increased assay numbers (**Fig.9A**). Similar results were obtained using shorter (0.5 min) or longer (8 min) annealing times. However, Cq values of template-containing samples did not decrease significantly between 3 and 8 min annealing time, implying that 3 min of annealing is sufficient for effective target binding under these conditions. Interestingly, not only the total PCR product formation increased with increasing assay numbers, but also the PCR product yield of individual assays in downstream RT-qPCR (**Fig.9B and C**), along with improved reproducibility. Due to the large total number of different primer pairs present in the highly multiplexed preamplification reaction, non-specific PCR products are formed at large quantities. One explanation for this observation may be that increased numbers of primer pairs during preamplification will increase the formation of possible primer-to-primer interactions as well as the formation of non-specific PCR products. However, nonspecific PCR products formed during preamplification will only interfere with the downstream singleplex PCR if the particular nonspecific PCR product is complementary to the applied primer pairs.
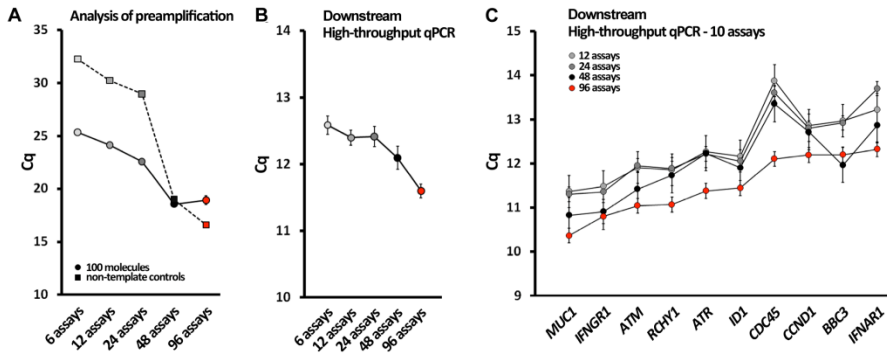
***Figure 9.   Assay number dependence. A.** Cq-values (average ±SD) for positive
(n=3) and negative samples (n=3) using different number of assays in
preamplification. **B.** High-throughput qPCR data of individual assays. Average
Cq ±SD (n=3) is shown. Data from all preamplified genes were used. **C.** Average
Cq ±SD (n=3) of 10 assays included in the preamplification with 12, 24, 48 and
96 pooled assays.*

Dependence on primer concentration, annealing time and temperature

Primer concentration, annealing time and duration of the annealing step
are reciprocal factors in preamplification. To reduce the formation of
nonspecific PCR product formation in the multiplex reaction primer
concentrations are 10-20 times lower compared to regular PCR [85, 109].
To maintain high preamplification efficiency at low primer concentration
the annealing time is usually extended up to several minutes. The effect of
variable primer concentrations (10, 40, 160, 240 nM) was tested in
relation to different annealing times (0.5, 3, 8 min). Analysis of
preamplification revealed elevated yields of specific and non-specific
PCR products as primer concentrations and annealing times were
increased. We observed a shift from specific towards non-specific product
formation when primer concentrations were increased from 40 to 160 nM.
The performance of individual assays was dependent on the primer
concentration and annealing time as well. We found individual assays
performed best at a concentration of larger than 40 nM using long
annealing times (3 min and 8 min).

All primers applied for this study were designed to anneal to their specific
target sequence at 60 °C. Using analysis of preamplification of an
annealing temperature gradient ranging between 55.0 °C to 65.3 °C we

made two main observations: First, an increase in annealing temperature lead to a reduction of PCR product yields. Second, we detected a gradual shift from non-specific to specific product formation as the annealing temperature increased. For downstream qPCR highest yield, specificity and reproducibility was observed at annealing temperatures between 58.5℃ and 61.3℃, using assays optimized to anneal at 60℃.

Effect of various PCR additives and single-cell gene expression profiling

Analysis of preamplification revealed large amounts of non-specific PCR products formed for most tested conditions. Therefore, we have tested the effects of 18 different PCR additives in 35 different reactions, which may improve enzymatic reactions involving nucleic acids. The formation of nonspecific PCR products was reduced by 10 cycles (~1000-fold) compared with preamplification without additives when using and 2 mg/mL bovine serum albumin supplied with 2.5 and 5.0% glycerol, respectively, 5%, glycerol, 0.5 M formamide and 0.5 M L-carnitine. The effect of nine selected additives was further evaluated at the individual assay level, using downstream qPCR of 96 assays. Here, the preamplification performed equally regardless whether additives were present or not. Most likely this is because our assays are extensively optimized for high efficiency, specificity and sensitivity. However, PCR additives may prove beneficial for less optimized assays or in the context of next-generation sequencing where formation of non-specific products may impede sequencing capacity and reduce the amount of informative reads.

Today, many clinical applications strive towards the use of non-invasive sampling strategies and small biopsies, including fine needle aspirates and circulating tumor cells, to detect and quantify biomarkers. Due to the low abundance of starting material, adequate molecule quantification requires highly sensitive, robust and specific technologies [110-112]. The preferred strategy to quantify multiple DNA or cDNA targets in biological samples of limited size is to first preamplify the material, which theoretically allows for the analysis of any target sequence by downstream qPCR or next generation sequencing. Optimized preamplification protocols typically show high sensitivity, specificity, efficiency, reproducibility and dynamic range. Targeted preamplification is usually conducted as a

multiplex PCR, restricting the amplification to the sequences of interest only [109, 113].

In conclusion our data suggests, that the number of preamplification cycles should be sufficient to produce at least five (accurate sensitivity), but preferentially 35 (accurate precision) molecules per downstream qPCR reaction. A small number of highly abundant targets will likely not affect the performance of other assays. Furthermore, we found that the usage of large assay pools, low primer concentration and long annealing times is beneficial for accurate targeted preamplification.

4.2 Results and discussion paper II

Breast cancer is a distinctly heterogeneous disease with respect to histological, molecular and clinical features, affecting disease progression and treatment response [114]. The cancer stem cell model may provide one explanation for the observed intratumoral heterogeneity, suggesting that cancers are driven by a cellular subpopulation with stem cell properties, which give rise to hierarchically structured tumors. Currently, there is a lack of universal and definite CSC markers, indicating that the CSC phenotype may not necessarily be uniform between cancer subtypes or even tumors of the same subtype [55]. Categorization of CSCs is further complicated by their cellular plasticity [50-53] and a dynamic microenvironment [39].

In paper II we aimed to characterize putative CSC pools in ERα+ and ERα- models of breast cancer. To this end, we established single-cell RT-qPCR-based gene expression profiling of well-known markers of differentiation, stemness, the EMT and cell cycle regulators. To circumvent current obstacles associated with immunophenotype-based CSC-enrichment methods, in this study we applied three functional *in vitro* CSC assays; growth in anchorage-independent culture, hypoxia and isolation of low proliferative, label-retaining cells derived from mammospheres (**Fig.10A-C**). All methods have previously been demonstrated to enrich for cells with increased cancer initiating potential in mouse model systems [60, 91, 115].
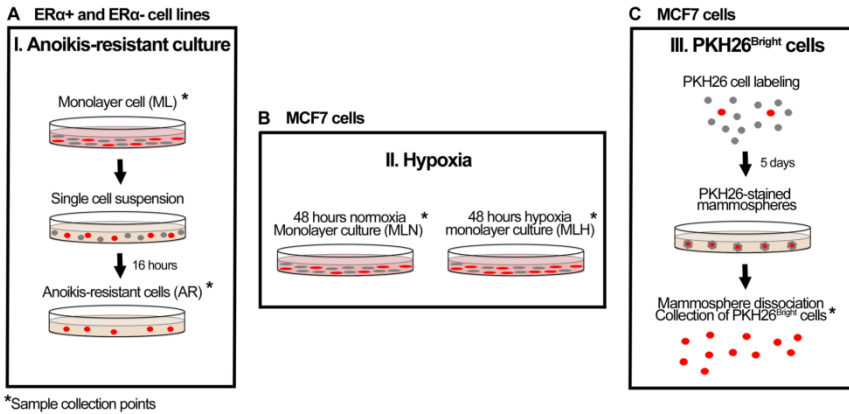
Figure 10. Applied functional CSC enrichment methods. Breast cancer cell lines were cultured as regular monolayers and cancer stem like cells were enriched using three established techniques: A. Growth in anchorage-independent culture (ERα+ and ERα- cell lines). B. Hypoxia (1% O_2 for 48 h) (MCF7 cells). C. Non-dividing, PKH26^{Bright} cells cultured as mammospheres (MCF7 cells).

ERα+ cell lines display distinct subpopulations with CSC-like and differentiated phenotypes, while proliferative phenotypes define ERα- breast cancer cell lines

To detail CSC in the first CSC enrichment approach, we made use of the ability of CSCs to withstand anoikis in anchorage-independent culture systems [88, 91]. CSC were enriched following 16-hours growth in anchorage-independent conditions and analyzed in parallel with matched monolayer cultures. To investigate cellular organization as well as the CSC pool in ERα+ and ERα- breast cancer models, individual ERα+ cells (MCF7, n=157; T47D, n=158) and ERα- cells (CAL120, n=140; MDA231, n=159) were subjected to single-cell gene expression profiling.

Using principal component analysis (PCA), monolayer and anoikis-resistant MCF7 cells displayed three distinct clusters, termed ERα+ I-III. ERα+ I was characterized by high expression of the pluripotency-associated genes, lack of proliferation markers and low overall expression

levels, characteristic for quiescent stem cells [116, 117]. ERα+ II exhibited high expression of breast cancer stem cell associated genes as well as high expression of the proliferation markers. ERα+ III was denoted by high expression of differentiation-associated genes. Anoikis-resistant cells were enriched in clusters ERα+ I and II, whereas the majority of monolayer cells was present in cluster ERα+ III. Similar clusters were observed for T47D cells; we identified two clusters ERα+ I and III. Interestingly, differential expressed genes between anoikis-resistant cells and monolayer cells were essentially identical for the two analyzed ERα+ cell lines, suggesting similar CSC enrichment mechanisms within this breast cancer subtype.

In line with previously published data, single-cell analysis has demonstrated that the majority of regular grown ERα+ cells displayed a RNA expression profile reminiscent of a more differentiated luminal phenotype [118, 119]. In contrast, ERα+ anoikis-resistant cells formed well-separated clusters with distinct CSC-like gene expression signatures, indicative of a hierarchical cell organization. Intriguingly, for MCF7 cells we have identified two clusters with distinct CSC-like gene expression profiles, which were enriched for anoikis resistant cells. This data points towards the presence of multiple CSC-like pools. Based on the observed gene expression profiles the two clusters could represent alternative CSC-like or differentiation states. Alternatively, differences in the transcriptomic phenotype may also result from cellular subpopulations featuring a distinct genetic/epigenetic background. As has been suggested, stochastic clonal evolution and the stem cell hypothesis are not mutually exclusive [120]. Using single-cell transplantation assays, two recent publications have described genetic diversity and clonal evolution of leukemic CSCs [121, 122]. Yet, the definite description of various CSC pools and their therapeutic relevance requires further functional characterization. In addition, to correlate genotypes with transcriptional or protein phenotypes, protocols for the detection of DNA, RNA and protein derived from the same cell have been described [123].

We next scrutinized two ERα- cell lines using the same experimental setup as for ERα+ cells. For CAL120 cells PCA identified two clusters, termed ERα- I and III, in accordance with the nomenclature used for ERα+ cells. ERα- I cells were characterized by low total RNA levels, while ERα- III were characterized by high expression of 14 genes, belonging to all defined gene groups. The majority of cells was present in cluster ERα- III, anoikis-resistant cells were slightly enriched in cluster ERα- I. MDA231 cells formed two clusters, termed ERα- II and III.

Comparison of differential gene expression between anoikis-resistant and monolayer cells revealed that most genes were down-regulated after 16-hours of anchorage-independent culture. As opposed to ERα+ cells, *ID1* and *CCNA2* were the only commonly down-regulated genes across the two cell lines, perhaps reflective of the heterogenetic nature of this breast cancer subgroup.

Compared to the ERα+ cell lines, the segregation of ERα- monolayer and anoikis-resistant cells was less pronounced. Separation into distinct clusters was mainly due to differences in their proliferative capacity (data not shown). The reasons for that could either be that our applied gene panel did not ideally separate CSC-enriched fractions from regular grown cells or that ERα- cell lines do not feature a strict hierarchical organization, in line with observations for melanomas [124]. ERα- monolayer and anoikis-resistant cells displayed a characteristic basal/mesenchymal phenotype [119], which may in part mask differentiation [103]. Our results nevertheless suggest that ERα- breast CSC cluster based on proliferative capacity. To further investigate the applicability of current CSC markers and to identify novel pathways specific to CSC in both luminal (ERα+) and basal (ERα-) breast cancer subtypes we have applied a RNA sequencing approach of CSC-enriched fractions in conjunction to matched monolayer cultures (see paper III).

A common quiescent CSC-like subpopulation can be identified in ERα+ and ERα- cell lines

To scrutinize the relationship between different breast cancer subtypes and the presence of CSC markers, we conducted combined multivariate analyses of all cells and grouped them by similarities in their gene expression profiles. Multiple clustering algorithms defined three discrete clusters for ERα+ cell lines (ERα+ I-III), whereas ERα- cell lines congregated into three partly separate clusters (ERα- I-III). ERα+/ERα- I cluster included cells of all cell lines. Cluster ERα+ II mainly contained MCF7 AR cells, whereas cluster ERα+ III encompassed the majority of all differentiated ERα+ ML cells. Clusters ERα- II-III harbored essentially all MDA231 cells as well as most of the CAL120 cells. The clusters defined low (ERα- II) or high (ERα- III) proliferative groups. The cellular organization of both ERα+ and ERα- cells is schematically illustrated in **Figure 11A**.

Comprehensive analysis of all cells revealed a clustering characteristic of hierarchical organization for the analyzed ERα+ cells. Furthermore, the data may suggest that MCF7 and T47D cells exhibit two separate modes of differentiation. MCF7 cells seemed to differentiate from a quiescent CSC-like cell state (ERα+/ERα- I) via a progenitor-like state (ERα+ II) to acquire a more differentiated phenotype (ERα+ III), while T47D cells did not seem to pass through this progenitor-like state. ERα- cell lines on the other hand were mainly separated by their increasing proliferative capacity from a common quiescent CSC-like pool, shared with ERα+ cells.

Our data indicates the presence of a quiescent CSC-like pool in both breast cancer subtypes, based on the expression of pluripotency-associated genes and low overall transcript levels, which has been described for cells in a dormant state [116, 117, 125]. Upon differentiation, ERα+ and ERα- cell lines activate partly different pathways by regulating specific genes which give rise to the more mature cell types that characterize these breast cancer subtypes.

To validate our findings in a clinical context we analyzed single-cells derived from two freshly dissociated primary ductal breast cancer samples, one ERα+ (n=81) and one ERα- (n=90). Combined PCA of the two tumors cells revealed a clustering pattern based on their origin (ERα+ or ERα-), but with an overlap of some cells sharing a similar gene expression profile. This common cell pool was characterized by the expression of pluripotency markers, while the other cells expressed markers related to more differentiated cell states. The number of cells with a common undifferentiated gene expression profile was rather high, potentially including both common progenitor cells as well as CSCs. **Figure 11B** illustrates the differentiation route in primary tumor cells, which was in line with the cell hierarchy delineated for cell lines. Further analysis, *e.g.* by using RNA sequencing of larger cell line and patient cohorts at the individual cell level will most likely reveal the presence of additional cellular subpopulations. From a therapeutic point of view the identification of different CSC is highly relevant. In order to design curative treatment approaches all tumor-propagating populations need to be eradicated.
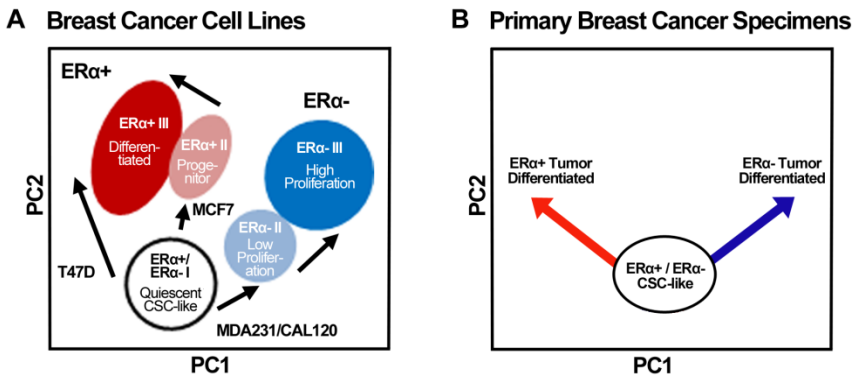
*Figure 11. ERα+ and ERα- cells define a common quiescent CSC pool. A: Hypothesized cellular organization of ERα+ and ERα- cell lines. **B:** Hypothesized cellular organization of ERα+ and ERα- primary tumors.*

ERα+ MCF7 cells comprise distinct cellular states and are organized in a hierarchical manner

Since the applied gene panel proved more suitable to detect cellular subpopulations in ERα+ cell lines, for succeeding experiments we continued with the ERα+ MCF7 cells. For a detailed investigation of CSC-like/progenitor pools we used two additional functional CSC enrichment approaches, namely 1% hypoxia (**Fig.10B**) and PKH26-label retention in anchorage-independent culture (**Fig.10C**) and conducted single-cell analysis. Combined PCA and Kohonen self organizing map (SOM) analyses of all enriched MCF7 CSC-fractions and matched monolayer cultures, allowed us to relate and organize phenotypic states. Using SOM individual cells established four stable clusters (MCF7 I-IV) based on differential transcriptomic profiles, schematically shown in **Figure 12**. Clusters MCF7 I-IV each contained cells from all applied enrichment methods, although in varying proportions. Cluster MCF7 I harbored mainly anoikis-resistant cells and displayed high expression of EMT-, pluripotency-, and certain breast cancer stem cell-related genes. Cluster MCF7 II primarily contained PKH26^Bright cells and was characterized by high expression of *CD44*. Cluster MCF7 III was enriched for hypoxic cells and to a lesser extent for PKH26^Bright cells with high expression of most differentiation markers as well as *ABCG2* and *ERBB2*. Most monolayer cells were present in cluster MCF7 IV characterized by

high expression of proliferation-associated genes, *PGR*, *ALDH1A3* and *ID1*.

The observed gradual gene regulation between the identified clusters suggests a hierarchical organization of MCF7 cells. The MCF7 I group features the phenotype of quiescent CSCs and represents the apex of the hierarchy and differentiation takes place over different cellular states (MCF7 II and MCF7 III) to the most differentiated cells in group MCF7 IV. First, differentiation-associated genes were activated in immature CSCs at the same time as EMT and breast cancer associated stem cell markers were downregulated. Secondly, we observed increased expression of proliferation markers and downregulation of genes related to stemness. This progression sequence is further in line with normal stem cell differentiation and development [126, 127].
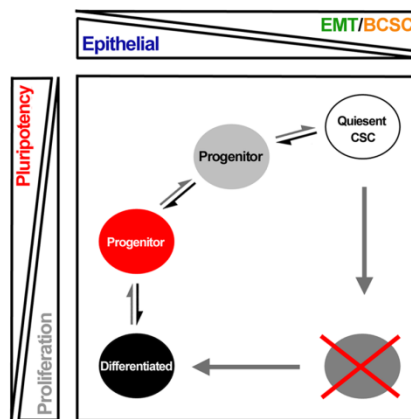

Figure 12. ERα+ MCF7 cells feature distinct differentiation states organized in a hierarchical manner. Proposed model displaying distinct identified cell states and hierarchical organization of MCF7 cells. The trend of gene expression of epithelial/differentiation, breast cancer stem cell (BCSC), pluripotency, EMT/metastasis and proliferation associated genes are indicated outside the box.

In conclusion, our data suggest that ERα+ and ERα- cell lines share a quiescent cell pool with CSC-like features. This phenotype was partly recapitulated in two primary tumor samples. Currently, it is not known, whether progenitor and CSC-like cells are similar across different molecular subtypes. The CSC concept comprises two separate

components; the first concerns the cell of origin of breast cancer and the second concerns the cell types responsible for tumor maintenance and progression [120]. Today it is widely believed that the different molecular subtypes arise from distinct cell types within the mammary hierarchy, but also particular oncogenic drivers seem to be involved in producing the various breast cancer phenotypes [39]. Basal (ERα-) cancers for example are thought to arise from a luminal progenitor cells [128]. The cellular origin of luminal cancers has yet to be established, however it has been speculated that a more differentiated luminal progenitor could give rise to this highly differentiated breast cancer type. In light of this it is possible that distinct subtypes harbor individual CSC-like/progenitor populations. Besides, CSCs in particular cells displaying the CD44[+] phenotype have been linked to the formation of metastasis [56]. Clinically, ERα+ and ERα- breast cancers show distinct organ-specific metastasis. ERα+ preferentially metastasize to the bone, while ERα- breast cancers tend to metastasis to visceral organs or to the brain [129]. This observation further underlines the possibility of distinct subtype-specific CSC/progenitor cells. On the other hand, although different subtypes exhibit a different mutational spectrum and the predominance of different cell types, it is possible, that CSCs depend on specific pathways, which may be shared across the molecular subtypes or even different cancer types. For example hedgehog signaling and the polycomb protein Bmi-1 have been demonstrated to regulate self-renewal in both, malignant and non-malignant stem cells of the breast [130]. Furthermore, a recent study has analyzed transcriptomic profiles of CSCs with the CD44[+]/CD24[-] and ALDH[+] phenotypes across different subtypes and found a remarkable similarity in CSC-derived gene expression patterns [61]. The identification of common gene-expression profiles in CSCs across molecular subtypes indicates that CSC-targeting agents could be effective in different types of breast cancer in combination with subtype-specific treatment [120]. One such an agent is the antibiotic salinomycin, which has been identified in a high-throughput screening [131] and proved effective in eradicating CSCs in multiple breast cancer types [132, 133].

The multiple CSC enrichment methods used to analyze MCF7 cells allowed for a detailed description of cell pools present at the individual cell level. We have identified four different populations, seemingly organized in a hierarchical manner as displayed by gradual up- and down regulation of differentiation-, stemness-, EMT-, and proliferation associated genes. Whether cells transition through multiple cellular differentiation states in a uni- or bidirectional manner has not explicitly

been addressed in this study, however several lines of evidence have recently reported a high degree of cellular plasticity and the capability of cells to switch between multiple cellular phenotypes [52, 53, 61, 134, 135]. Moreover, mathematical modeling has demonstrated that higher levels of dedifferentiation reduce the effect of CSC-targeted therapy and lead to higher rates of resistance [136], laying emphasis on the importance to take cellular plasticity into account when designing new treatment approaches. Our data permit the identification of key events in CSC plasticity. For example, in an attempt to target de-differentiation of progenitor cells into less-differentiated cells with pluripotent features, in ERα+ breast cancers, genes associated with differentiation/EMT/breast cancer stemness need to be modulated rather than pluripotency/proliferation, since these processes follow a sequential order. However, in ERα- cells, proliferation seems to be one of the key differentiation associated events. Targeting proliferation in both ERα+ and especially ERα- breast cancer may actually have an effect on differentiation processes potentially increasing CSC subpopulations and tumor aggressiveness. Our data highlight the need for proper tumor characterization and in depth understanding of relevant common as well as separate differentiation and de-differentiation processes present in subtypes of breast cancer.

4.3 Results and discussion paper III

In paper III we sought to identify molecular pathways specific to luminal and basal breast cancer subtypes, using cell line models. To this end, we conducted RNA sequencing of functionally enriched CSC fractions and matched monolayer cultures, to identify circuits commonly overrepresented in CSC fractions of both subtypes.

The mevalonate pathway is a key feature of CSC-enrichment in luminal and basal breast cancer subtypes

To identify commonly upregulated transcriptional networks in CSCs of luminal and basal subtypes of breast cancers, we applied a genome-wide RNA-sequencing approach. For this purpose CSC of luminal and basal breast cancer cell lines were enriched using a 16-hour anchorage-independent culture system. CSC-derived gene expression signatures were

analyzed in conjunction with corresponding adherent populations. A total of 344 (MCF7), 243 (T47D) and 477 (MDA231) genes were significantly upregulated in CSC-enriched subpopulations (**Fig.13A**). Ingenuity® Pathway Analysis (IPA®) identified mevalonate-associated networks in six out of the twelve most over-represented pathways (**Fig.13B**). Therefore, we have analyzed the expression of 11 mevalonate genes using single-cell RT-qPCR. Across the three cell lines, we identified similar expression patterns for *HMGCS1*, *MVK* and *DHCR24*. *HMGCS1* expression showed no significant difference between cell lines and was specific to a tumor cell sub-fraction. Following CSC enrichment, the percentage of *HMGCS1*-expressing cells was significantly expanded as analyzed at the individual cell level. *HMGCS1* did not exhibit any gene associations to other pathway genes, which may indicate mevalonate-pathway independent transcriptional regulation. Furthermore, *HMGCS1* was significantly over-expressed in all three CSC-enriched cell lines in the original RNA-sequencing dataset. This observation was confirmed with RT-qPCR, using RNA extracted from cell lines (**Fig.13C**). A recent study recognized the mevalonate pathway overrepresented in CSCs of basal cancers [75]. In this study gene expression signatures were derived from mammosphere cultures, containing mammosphere-initiating and differentiated progenitor cells [137], possibly masking pathways active in CSCs. By using 16-hours suspension culture we have previously shown to enrich for cells with increased *in vitro* mammosphere formation capacity, *in vivo* tumor formation as well as for cells with elevated expression of the CSC-like EPCAM^{+}/CD44^{+}CD24^{low} immunophenotype [91]. Based on our findings we further explored the putative role of HMGCS1 as a marker of functionally enriched CSCs.
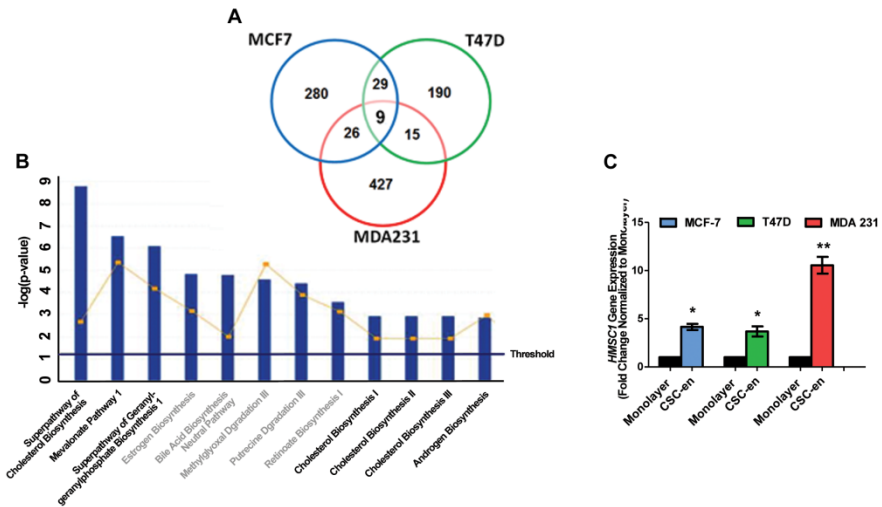
Figure 13. The mevalonate pathway is a key feature of CSC-enrichment in luminal and basal breast cancer subtypes. A. *RNA-sequencing was conducted on MCF-7, T47D and MDA231 adherent monolayers and 16-hour CSC-enriched cultures. Overall analysis of the RNA-sequencing data identified 344 (MCF-7), 243 (T47D) and 477 (MDA231) genes significantly overexpressed in the CSC-enriched cultures compared to adherent monolayer cultures.* **B.** *Ingenuity® Pathway Analysis was applied to the 79 genes which were significantly overexpressed in two or more of the CSC-enriched cell line subpopulations. Several mevalonate pathway-associated networks were significantly increased above the statistical threshold (p<0.001) in CSC-enriched subpopulations.* **C.** *MCF-7, T47D and MDA-231 cell lines grown as adherent monolayer or in 16-hour CSC-enriched (CSC-en) cultures; HMGCS1 gene expression was assessed by qPCR. HMGCS1 transcript levels are reported as fold changes relative to individual monolayer controls, graphs represent average ±SEM, n=3.*

HMGCS1 is required for CSC survival in luminal and basal cell lines with mutated p53

HMGCS1 protein expression was studied in 16-hour anchorage-independent culture and in 5-day mammosphere cultures. Protein expression was increased in both model systems compared to adherent monolayer cultures, confirming the RNA data. To assess the effect of HMGCS in CSC function, gene expression was transiently silenced, using siRNA (**Fig.14A**) and survival in 16-hours and 5-days anchorage-independent cultures was analyzed. Following *HMGCS1* knockdown T47D and MDA-231 cells exhibited a significant decrease in survival

(**Fig.14B**) and relative mammosphere forming capacity (**Fig.14C**). Dysregulated mevalonate pathway activation has previously been linked to a gain-of-unction mutation of p53 [74]. The authors identified functional interaction with sterol regulatory element-binding proteins to be critical for mutant p53-mediated upregulation of mevalonate pathway genes.

To block the mevalonate pathway cells were treated with standard doses of simvastatin, targeting 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), the major rate-limiting enzyme within the mevalonate pathway and downstream substrate of HMGCS1 [138]. Simvastatin treatment induced elevated HMGCS1 protein levels, possibly through a restorative feedback response [139]. This feedback mechanism has been associated with statin-insensitivity and the lack to induce apoptosis upon statin exposure in leukemic cells [72]. Simvastatin effects on viability in 16-hour CSC-enrichment and mammosphere formation were negligible. Given the pronounced effect of HMGCS1 gene-silencing on CSC survival and seemingly mevalonate-pathway-independent regulation of HMGCS1, pharmacological inhibition of HMGCS1 may be a superior approach to specifically target CSC in luminal and basal breast cancer subtypes, however this requires further investigation.
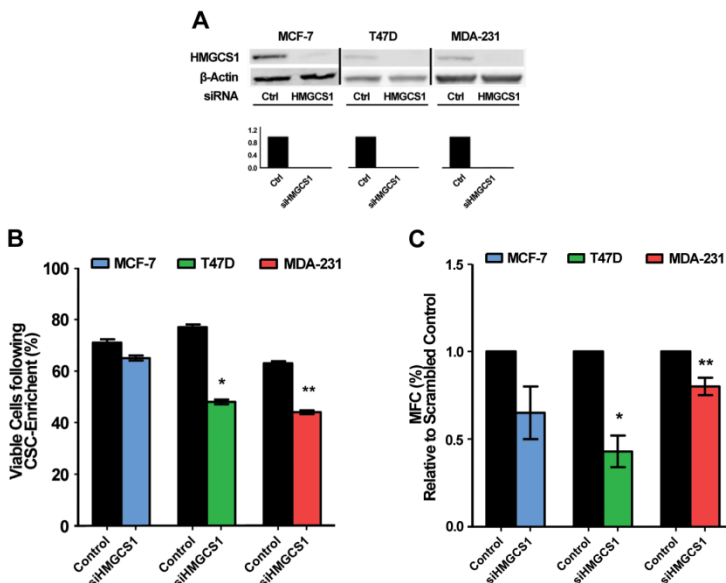
Figure 14. HMGCS1 is required for CSC survival in luminal and basal cell lines with mutated p53. A. HMGCS1 gene expression was silenced using HMGCS1 siRNA or scrambled negative control. Immunoblot and associated densitometry confirmed HMGCS1 knockdown relative to scrambled control. β-actin was used as loading control. B. Cell lines were transfected with HMGCS1 siRNA or scrambled negative control for 72 hours. Cells were grown in 16-hour suspension culture. Manual viability counts were conducted with Trypan Blue exclusion dye. Percentage viable cells of total cell number seeded are shown, graphs represent average counts ±SEM, n=3. C. Cell lines were transfected with HMGCS1 siRNA or scrambled negative control for 72 hours; cells were grown in 5-day mammosphere cultures. Mammosphere forming capacity (MFC) was assessed and is represented as percentage of total cells seeded. Graphs show average fold-change relative to scrambled control ±SEM, n=3.

HMGCS1 expression transcriptionally regulates CSC-associated genes in luminal and basal models of breast cancer with mutated p53

To investigate the gene regulatory role of HMGCS1, transcript levels of proliferation, pluripotency, EMT and breast cancer stem cell genes were assessed in *HMGCS1*-knockdown cells and scrambled controls. Following *HMGCS1* knockdown, T47D cells displayed a reduction of *ABCG2* expression, while MDA231 cells showed a decrease of *SOX2*, *NANOG* and *SNAI1* expression. No significant differences were detected in MCF7 cells, further suggesting a link between p53 mutational status and deregulated mevalonate metabolism. The adherent monolayer single-cell data was stratified by *HMGCS1* expression. *HMGCS1*-expressing cells had remarkably similar frequency in each cell line, however due to the low number of *HMGSC1*-expressing cells it was difficult to detect significant differences.

HMGCS1 expression correlates with disease aggression and associates with basal tumors in a breast cancer patient cohort

HMGCS1 expression was assessed in a cohort of 149 lymph node positive breast cancers (**Fig.15A**). Correlation analysis identified significant positive associations with p53 mutational status, tumor grade and HIF 1-α. HMGCS1 exhibited significant inverse correlations with ERα and PR, there was an obvious increase in the frequency of ERα- tumors in the HMGCS1-intermediate and -high tumors (**Fig.15B**) as well as an increase in p53 mutations in ERα+ tumors (**Fig.15C**), however due to the small

number of patients included in each HMGCS1 subgroup, these results are only indicative.
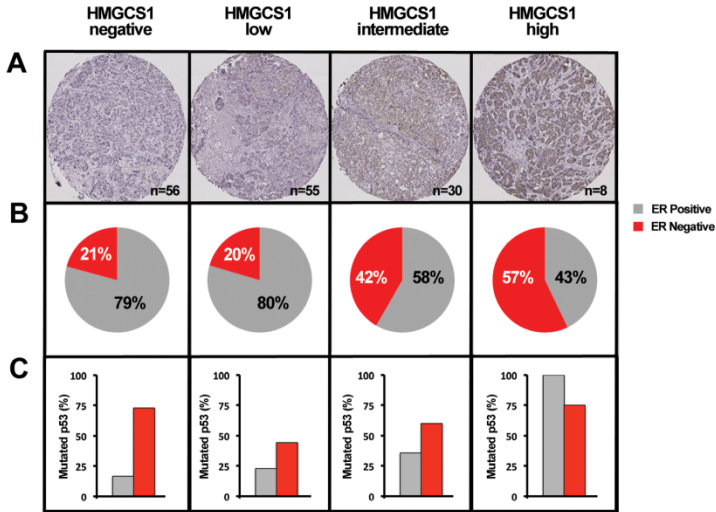

Figure 15. HMGCS1 expression correlates with disease aggression and associates with basal tumors in a breast cancer patient cohort. A. A tissue microarray of 149 tumor biopsies from patients with lymph-node positive breast cancers was stained for HMGCS1. Expression was assessed using the Allred scoring system. B. Comparative pie charts were generated for each HMGCS1 subgroup (negative, low, intermediate and high) to demonstrate the frequency of ERα+ and ERα- tumor subtypes present in each. C. Percentage of p53 mutations in ERα+ and ERα- breast cancers for HMGCS1 subgroups.

Hyperactivation of the mevalonate pathway and overexpression of mevalonate pathway genes has previously been linked to p53 gain-of-function mutations [74]. Here, we establish a specific link to HMGCS1 protein expression and mutated p53. Furthermore, HMGCS1 was associated with tumor grade and lack of ERα expression. About 30% of all breast cancers exhibit mutations in the *TP53* gene, but the frequency differs greatly across distinct molecular subtypes [23, 35]. About 80% of the basal tumors display mutations in the *TP53* gene, whereas only about 15% of luminal A tumors carry a p53 mutation [23, 35]. Furthermore,

basal tumors are characterized by the lack of ERα, PR and HER2 expression as well as high tumor grade [140]. Our data suggests that that therapeutic modulation of the mevalonate pathway may be beneficial for patients presenting p53 gain-of-function mutations and consequential mevalonate pathway hyperactivation, independent of breast cancer subtype.

In conclusion, by applying genome-wide RNA sequencing of functionally enriched CSCs in conjunction with matched adherent cultures, we identified the mevalonate pathway for cholesterol biogenesis and protein prenylation in CSCs of luminal and basal breast cancers. In breast cancer the mevalonate pathway has been implicated in tumor cell transformation, malignancy and the specific regulation of basal-derived CSCs [72, 74, 75]. Dissecting the mevalonate pathway, we identified HMGCS1 as a functional marker for CSCs. Transient HMGCS1 gene-silencing lead to reduced survival of CSCs in cell line model displaying mutated p53. Given the pronounced effect of HMGCS1 gene silencing on CSC survival pharmacological inhibition of HMGCS1 may be a promising approach to specifically target CSC in luminal and basal breast cancer subtypes and likely additional breast cancer types.

5 CONCLUSIONS

In **paper I** we have evaluated experimental parameters of targeted preamplification for samples with minute starting amount, such as single-cell measurements. Quantification of small molecule numbers often requires a preamplification step to produce sufficient material for accurate downstream applications. Based on our findings we were able to provide general recommendations for conducting robust and accurate preamplification in conjunction with RT-qPCR or next-generation sequencing.

In **paper II** we have evaluated the cellular organization of ERα+ and ERα- breast cancer cells by combining single-cell gene expression profiling with three functional enrichment techniques for cancer stem cell enrichment. ERα+ cells displayed a hierarchical organization with different modes of cellular transitions. ERα- cells clustered mainly based on their proliferative capacity. Despite of the seemingly different cellular organization of both subtypes, we have identified a quiescent cell pool with CSC-like features common to ERα+ and ERα- cells. In addition, for ERα+ MCF7 cells we identified several cellular populations, which displayed distinct CSC-like phenotypes, underlining the importance for detailed tumor characterization in order to target all therapeutic relevant cancer cells.

In **paper III** we have evaluated transcriptional networks in CSC-enriched populations of ERα+ and ERα- breast cancer cells and thereby identified the mevalonate pathway commonly overrepresented in both breast cancer subtypes. Detailing the mevalonate pathway we identified the mevalonate precursor enzyme HMGCS1 as a specific marker for CSC enrichment, essential for CSC activities, in particular in cells with mutated p53. Pharmacological inhibition of HMGCS1 could therefore be a promising new treatment approach to target CSCs in an appropriate p53 mutational background.

ACKNOWLEDGEMENT

First and foremost I would like to express my deep gratitude to my main supervisor **Prof. Göran Landberg** for guiding me through my PhD experience and providing such an excellent and exciting research environment!

I would also like to express my gratitude to my co-supervisor **Prof. Anders Ståhlberg**, whose dedication to science is truly inspiring! Thanks for sharing your expertise, excitement end enthusiasm about all qPCR-related matters.

I am deeply grateful for have been given the opportunity to work together with all the brilliant and talented post-docs in the lab as well as for constant support, help and scientific inspiration. I have learned so much from you during the last four years. In no particular order I would like to thank **Paul Fitzpatrick** (special thanks for everything, starting in San Fran, ending in the same office), **Eva Bom** (special thanks for FACSing my cells so skillfully and your valuable project input), **Susann Busch** (special thanks for proof-reading my thesis and your smart brain), **Hanna Jacobsson** (special thanks for proof-reading my thesis, Swedish translations and an always open ear), **Claire Walsh** (special thanks for being able to work on your project and very special birthday songs!), **Éamon Hughes** (special thanks for always being so helpful. Be assured the present thesis is a 100% cabbage-free.) as well as **Daniel Andersson** (special thanks for late night single-cell picking sessions and extreme qPCRing). It was such a pleasure working together with you guys!

I would also like to thank all the wonderful people in my group I have been working alongside with during the last couple of years. The epicenter of the lab **Pernilla Gregersson** (special thanks for advanced single-cell picking), **Ylva Magnusson**, **Pauline Isakson**, and **Svanheidur Rafnsdottir**.

In addition, I would like to offer my thanks to a bunch of people loosely associated to the lab or my project. **Isabell Lebküchner** (special thanks for endless single-cell data discussions and horse-riding adventures), **David Svec** (special thanks for qPCR expertise and Czech honey), **Soheila Dolatabadi** (special thanks for just the right amount of irony) and last but not least **Emma Jonasson** (special thanks for last-minute-Friday-afternoon-Swedish-abstract-writing!).

Meine Familie und Freunde vielen Dank für all Eure Unterstützung und Liebe.

Karl my love, the man who was brave enough to put up with crazy woman for the last few months!

REFERENCES

1.   Oakes SR, Gallego-Ortega D, Ormandy CJ. The mammary cellular hierarchy and breast cancer. **Cellular and Molecular Life Sciences**. 2014;71:4301-4324.
2.   Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. **Gene Dev**. 2009;23:2563-2577.
3.   Visvader JE, Stingl J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. **Gene Dev**. 2014;28:1143-1158.
4.   Lim E, Vaillant F, Wu D et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. **Nature Medicine**. 2009;15:907-913.
5.   Stingl J, Eaves CJ, Kuusk U et al. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. **Differentiation**. 1998;63:201-213.
6.   Keller PJ, Arendt LM, Skibinski A et al. Defining the cellular precursors to human breast cancer. **P Natl Acad Sci USA**. 2012;109:2772-2777.
7.   Ginestier C, Hur MH, Charafe-Jauffret E et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. **Cell Stem Cell**. 2007;1:555-567.
8.   Pece S, Tosoni D, Confalonieri S et al. Biological and Molecular Heterogeneity of Breast Cancers Correlates with Their Cancer Stem Cell Content. **Cell**. 2010;140:62-73.
9.   Sleeman KE, Kendrick H, Robertson D et al. Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. **Journal of Cell Biology**. 2007;176:19-26.
10.   Stingl J, Eaves CJ, Zandieh I et al. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. **Breast Cancer Res Tr**. 2001;67:93-109.
11.   Ferlay J, Soerjomataram I, Dikshit R et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. **Int J Cancer**. 2015;136:E359-386.
12.   Society AC. Breast Cancer Facts & Figures 2013-2014. Atlanta: American Cancer Society, Inc. 2013.
13.   Socialstyrelsen. Cancer incidence in Sweden 2011 **ISBN 978-91-7555-003-9**. 2012.
14.   Nathanson KN, Wooster R, Weber BL. Breast cancer genetics: What we know and what we need. **Nature Medicine**. 2001;7:552-556.

15.   Weigelt B, Geyer FC, Reis JS. Histological types of breast cancer:
       How special are they? **Molecular oncology**. 2010;4:192-208.
16.   Lee AHS, Ellis IO. The Nottingham prognostic index for invasive
       carcinoma of the breast. **Pathology & Oncology Research**.
       2008;14:113-115.
17.   Haybittle JL, Blamey RW, Elston CW et al. A Prognostic Index in
       Primary Breast-Cancer. **Brit J Cancer**. 1982;45:361-366.
18.   Rakha EA, Reis JS, Baehner F et al. Breast cancer prognostic
       classification in the molecular era: the role of histological grade.
       Breast Cancer Research. 2010;12.
19.   Osborne CK. Steroid hormone receptors in breast cancer
       management. **Breast Cancer Res Tr**. 1998;51:227-238.
20.   Payne SJL, Bowen RL, Jones JL et al. Predictive markers in
       breast cancer - the present. **Histopathology**. 2008;52:82-90.
21.   Gruvberger S, Ringner M, Chen YD et al. Estrogen receptor status
       in breast cancer is associated with remarkably distinct gene
       expression patterns. **Cancer Research**. 2001;61:5979-5984.
22.   Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human
       breast tumours. **Nature**. 2000;406:747-752.
23.   Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns
       of breast carcinomas distinguish tumor subclasses with clinical
       implications. **P Natl Acad Sci USA**. 2001;98:10869-10874.
24.   Reis JS, Pusztai L. Breast Cancer 2 Gene expression profiling in
       breast cancer: classification, prognostication, and prediction.
       Lancet. 2011;378:1812-1823.
25.   Herschkowitz JI, Simin K, Weigman VJ et al. Identification of
       conserved gene expression features between murine mammary
       carcinoma models and human breast tumors. **Genome Biol**.
       2007;8.
26.   Hallett RM, Dvorkin-Gheva A, Bane A et al. A Gene Signature
       for Predicting Outcome in Patients with Basal-like Breast Cancer.
       Sci Rep-Uk. 2012;2.
27.   Miller E, Lee HJ, Lulla A et al. Current treatment of early breast
       cancer: adjuvant and neoadjuvant therapy. **F1000Res**. 2014;3:198.
28.   Carlson RW, Hudis CA, Pritchard KI et al. Adjuvant endocrine
       therapy in hormone receptor-positive postmenopausal breast
       cancer: evolution of NCCN, ASCO, and St Gallen
       recommendations. **J Natl Compr Canc Netw**. 2006;4:971-979.
29.   Vu T, Claret FX. Trastuzumab: updated mechanisms of action and
       resistance in breast cancer. **Front Oncol**. 2012;2:62.
30.   Cho SH, Jeon J, Kim SI. Personalized medicine in breast cancer: a
       systematic review. **J Breast Cancer**. 2012;15:265-272.
31.   De Abreu FB, Schwartz GN, Wells WA et al. Personalized
       therapy for breast cancer. **Clinical genetics**. 2014;86:62-67.

32. Polyak K. Heterogeneity in breast cancer. **J Clin Invest**. 2011;121:3786-3788.
33. Skibinski A, Kuperwasser C. The origin of breast tumor heterogeneity. **Oncogene**. 2015;34:5309-5316.
34. Banerji S, Cibulskis K, Rangel-Escareno C et al. Sequence analysis of mutations and translocations across breast cancer subtypes. **Nature**. 2012;486:405-409.
35. Koboldt DC, Fulton RS, McLellan MD et al. Comprehensive molecular portraits of human breast tumours. **Nature**. 2012;490:61-70.
36. Visvader JE. Cells of origin in cancer. **Nature**. 2011;469:314-322.
37. Chaffer CL, Weinberg RA. Cancer Cell of Origin: Spotlight on Luminal Progenitors. **Cell Stem Cell**. 2010;7:271-272.
38. Campbell LL, Polyak K. Breast tumor heterogeneity: cancer stem cells or clonal evolution? **Cell Cycle**. 2007;6:2332-2338.
39. Marjanovic ND, Weinberg RA, Chaffer CL. Cell Plasticity and Heterogeneity in Cancer. **Clinical Chemistry**. 2013;59:168-179.
40. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? **Nature Reviews Cancer**. 2012;12:323-334.
41. Dalerba P, Kalisky T, Sahoo D et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. **Nature Biotechnology**. 2011;29:1120-U1111.
42. Shipitsin M, Campbell LL, Argani P et al. Molecular definition of breast tumor heterogeneity. **Cancer Cell**. 2007;11:259-273.
43. Clevers H. The cancer stem cell: premises, promises and challenges. **Nature Medicine**. 2011;17:313-319.
44. Kandoth C, McLellan MD, Vandin F et al. Mutational landscape and significance across 12 major cancer types. **Nature**. 2013;502:333-339.
45. Navin N, Krasnitz A, Rodgers L et al. Inferring tumor progression from genomic heterogeneity. **Genome Research**. 2010;20:68-80.
46. Visvader JE, Lindeman GJ. Cancer Stem Cells: Current Status and Evolving Complexities. **Cell Stem Cell**. 2012;10:717-728.
47. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. **Nature Medicine**. 1997;3:730-737.
48. Al-Hajj M, Wicha MS, Benito-Hernandez A et al. Prospective identification of tumorigenic breast cancer cells. **Proc Natl Acad Sci U S A**. 2003;100:3983-3988.
49. Rosen JM, Jordan CT. The Increasing Complexity of the Cancer Stem Cell Paradigm. **Science**. 2009;324:1670-1673.

50.   Mani SA, Guo W, Liao MJ et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. **Cell**. 2008;133:704-715.
51.   Morel AP, Lievre M, Thomas C et al. Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. **Plos One**. 2008;3.
52.   Chaffer CL, Brueckmann I, Scheel C et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. **Proc Natl Acad Sci U S A**. 2011;108:7950-7955.
53.   Gupta PB, Fillmore CM, Jiang G et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. **Cell**. 2011;146:633-644.
54.   Malhotra GK, Zhao X, Band H et al. Shared signaling pathways in normal and breast cancer stem cells. **J Carcinog**. 2011;10:38.
55.   Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. **Nature Reviews Cancer**. 2008;8:755-768.
56.   Liu H, Patel MR, Prescher JA et al. Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. **Proc Natl Acad Sci U S A**. 2010;107:18115-18120.
57.   Li X, Lewis MT, Huang J et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. **J Natl Cancer Inst**. 2008;100:672-679.
58.   Yu F, Yao H, Zhu P et al. let-7 regulates self renewal and tumorigenicity of breast cancer cells. **Cell**. 2007;131:1109-1123.
59.   Zhang M, Atkinson RL, Rosen JM. Selective targeting of radiation-resistant tumor-initiating cells. **Proc Natl Acad Sci U S A**. 2010;107:3522-3527.
60.   Harrison H, Rogerson L, Gregson HJ et al. Contrasting hypoxic effects on breast cancer stem cell hierarchy is dependent on ER-alpha status. **Cancer Res**. 2013;73:1420-1433.
61.   Liu S, Cong Y, Wang D et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. **Stem Cell Reports**. 2014;2:78-91.
62.   Hanahan D, Weinberg RA. The hallmarks of cancer. **Cell**. 2000;100:57-70.
63.   Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. **Cell**. 2011;144:646-674.
64.   Warburg O. On the origin of cancer cells. **Science**. 1956;123:309-314.
65.   Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. **Cell**. 2008;134:703-707.

66. Wiesener MS, Munchenhagen PM, Berger I et al. Constitutive activation of hypoxia-inducible genes related to overexpression of hypoxia-inducible factor-1 alpha in clear cell renal carcinomas. **Cancer Research**. 2001;61:5215-5222.

67. Zhong H, De Marzo AM, Laughner E et al. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. **Cancer Res**. 1999;59:5830-5835.

68. Kim JW, Dang CV. Cancer's molecular sweet tooth and the Warburg effect. **Cancer Research**. 2006;66:8927-8930.

69. Heiden MGV, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. **Science**. 2009;324:1029-1033.

70. Goldstein JL, Brown MS. Regulation of the Mevalonate Pathway. **Nature**. 1990;343:425-430.

71. Thurnher M, Gruenbacher G, Nussbaumer O. Regulation of mevalonate metabolism in cancer and immune cells. **Bba-Mol Cell Biol L**. 2013;1831:1009-1015.

72. Clendening JW, Pandyra A, Li Z et al. Exploiting the mevalonate pathway to distinguish statin-sensitive multiple myeloma. **Blood**. 2010;115:4787-4797.

73. Jiang P, Mukthavaram R, Chao Y et al. In vitro and in vivo anticancer effects of mevalonate pathway modulation on human cancer cells. **Brit J Cancer**. 2014;111:1562-1571.

74. Freed-Pastor WA, Mizuno H, Zhao X et al. Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway. **Cell**. 2012;148:244-258.

75. Ginestier C, Monville F, Wicinski J et al. Mevalonate metabolism regulates Basal breast cancer stem cells and is a potential therapeutic target. **Stem Cells**. 2012;30:1327-1337.

76. Gruenbacher G, Thurnher M. Mevalonate metabolism in cancer. **Cancer Letters**. 2015;356:192-196.

77. McTaggart SJ. Isoprenylated proteins. **Cellular and Molecular Life Sciences**. 2006;63:255-267.

78. Liu J, Zhang C, Hu WW et al. Tumor suppressor p53 and its mutants in cancer metabolism. **Cancer Letters**. 2015;356:197-203.

79. Levine AJ, Oren M. The first 30 years of p53: growing ever more complex. **Nature Reviews Cancer**. 2009;9:749-758.

80. Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. **Cell**. 2009;137:413-431.

81. Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. **Gene Dev**. 2012;26:1268-1286.

82. Muller PAJ, Vousden KH. p53 mutations in cancer. **Nat Cell Biol**. 2013;15:2-8.

83.   Muller PAJ, Vousden KH. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. **Cancer Cell**. 2014;25:304-317.

84.   Stahlberg A, Bengtsson M. Single-cell gene expression profiling using reverse transcription quantitative real-time PCR. **Methods**. 2010;50:282-288.

85.   Stahlberg A, Kubista M. The workflow of single-cell expression profiling using quantitative real-time PCR. **Expert Rev Mol Diagn**. 2014;14:323-331.

86.   Charafe-Jauffret E, Ginestier C, Birnbaum D. Breast cancer stem cells: tools and models to rely on. **Bmc Cancer**. 2009;9.

87.   Liu Y, Nenutil R, Appleyard MV et al. Lack of correlation of stem cell markers in breast cancer stem cells. **Br J Cancer**. 2014;110:2063-2071.

88.   Dontu G, Abdallah WM, Foley JM et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. **Genes Dev**. 2003;17:1253-1270.

89.   Farnie G, Clarke RB, Spence K et al. Novel cell culture technique for primary ductal carcinoma in situ: Role of notch and epidermal growth factor receptor signaling pathways. **J Natl Cancer I**. 2007;99:616-627.

90.   Ponti D, Costa A, Zaffaroni N et al. Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. **Cancer Research**. 2005;65:5506-5511.

91.   Harrison H, Farnie G, Howell SJ et al. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. **Cancer Res**. 2010;70:709-718.

92.   Favaro E, Lord S, Harris AL et al. Gene expression and hypoxia in breast cancer. **Genome Medicine**. 2011;3.

93.   Denko NC, Fontana LA, Hudson KM et al. Investigating hypoxic tumor physiology through gene expression patterns. **Oncogene**. 2003;22:5907-5914.

94.   Bos R, van Diest PJ, van der Groep P et al. Expression of hypoxia-inducible factor-1 alpha and cell cycle proteins in invasive breast cancer are estrogen receptor related. **Breast Cancer Research**. 2004;6:R450-R459.

95.   Coradini D, Pellizzaro C, Speranza A et al. Hypoxia and estrogen receptor profile influence the responsiveness of human breast cancer cells to estradiol and antiestrogens. **Cellular and Molecular Life Sciences**. 2004;61:76-82.

96.   Kimbro KS, Simons JW. Hypoxia-inducible factor-1 in human breast and prostate cancer. **Endocr-Relat Cancer**. 2006;13:739-749.

97. Cooper C, Liu GY, Niu YL et al. Intermittent hypoxia induces proteasome-dependent down-regulation of estrogen receptor alpha in human breast carcinoma. **Clinical Cancer Research**. 2004;10:8720-8727.

98. Ryu K, Park C, Lee Y. Hypoxia-inducible factor 1 alpha represses the transcription of the estrogen receptor alpha gene in human breast cancer cells. **Biochem Bioph Res Co**. 2011;407:831-836.

99. Jogi A, Ora I, Nilsson H et al. Hypoxia alters gene expression in human neuroblastoma cells toward an immature and neural crest-like phenotype. **P Natl Acad Sci USA**. 2002;99:7021-7026.

100. McCord AM, Jamal M, Shankavaram UT et al. Physiologic oxygen concentration enhances the stem-like properties of CD133+ human glioblastoma cells in vitro (vol 7, pg 489, 2009). **Molecular Cancer Research**. 2009;7:987-987.

101. Soeda A, Park M, Lee D et al. Hypoxia promotes expansion of the CD133-positive glioma stem cells through activation of HIF-1 alpha. **Oncogene**. 2009;28:3949-3959.

102. Schillert A, Trumpp A, Sprick MR. Label retaining cells in cancer - The dormant root of evil? **Cancer Letters**. 2013;341:73-79.

103. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. **Breast Cancer Res**. 2008;10:R25.

104. Moore N, Lyle S. Quiescent, slow-cycling stem cell populations in cancer: a review of the evidence and discussion of significance. **J Oncol**. 2011;2011.

105. Eyler CE, Rich JN. Survival of the fittest: Cancer stem cells in therapeutic resistance and angiogenesis. **Journal of Clinical Oncology**. 2008;26:2839-2845.

106. Gao MQ, Choi YP, Kang S et al. CD24(+) cells from hierarchically organized ovarian cancer are enriched in cancer stem cells. **Oncogene**. 2010;29:2672-2680.

107. Eberwine J, Yeh H, Miyashiro K et al. Analysis of gene expression in single live neurons. **Proc Natl Acad Sci U S A**. 1992;89:3010-3014.

108. Lao K, Xu NL, Sun YA et al. Real time PCR profiling of 330 human micro-RNAs. **Biotechnol J**. 2007;2:33-35.

109. Livak KJ, Wills QF, Tipping AJ et al. Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. **Methods**. 2013;59:71-79.

110. Alix-Panabieres C, Pantel K. OPINION Challenges in circulating tumour cell research. **Nature Reviews Cancer**. 2014;14:623-631.

111. Bettegowda C, Sausen M, Leary R et al. Detection of Circulating Tumor DNA in Early and Late Stage Human Malignancies. **Neuro-Oncology**. 2014;16.

112.   Pantel K, Alix-Panabieres C. Real-time Liquid Biopsy in Cancer Patients: Fact or Fiction? **Cancer Research**. 2013;73:6384-6388.

113.   Rusnakova V, Honsa P, Dzamba D et al. Heterogeneity of Astrocytes: From Development to Injury - Single Cell Gene Expression. **Plos One**. 2013;8.

114.   Bertos NR, Park M. Breast cancer - one term, many entities? **J Clin Invest**. 2011;121:3789-3796.

115.   Richichi C, Brescia P, Alberizzi V et al. Marker-independent method for isolating slow-dividing cancer stem cells in human glioblastoma. **Neoplasia**. 2013;15:840-847.

116.   Fukada S, Uezumi A, Ikemoto M et al. Molecular signature of quiescent satellite cells in adult skeletal muscle. **Stem Cells**. 2007;25:2448-2459.

117.   Huttmann A, Liu SL, Boyd AW et al. Functional heterogeneity within rhodamine123(lo) Hoechst33342(lo/sp) primitive hemopoietic stem cells revealed by pyronin Y. **Experimental hematology**. 2001;29:1109-1116.

118.   Charafe-Jauffret E, Ginestier C, Monville F et al. Gene expression profiling of breast cell lines identifies potential new basal markers. **Oncogene**. 2006;25:2273-2284.

119.   Neve RM, Chin K, Fridlyand J et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. **Cancer Cell**. 2006;10:515-527.

120.   Brooks MD, Burness ML, Wicha MS. Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. **Cell Stem Cell**. 2015;17:260-271.

121.   Anderson K, Lutz C, van Delft FW et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. **Nature**. 2011;469:356-361.

122.   Notta F, Mullighan CG, Wang JCY et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells (vol 469, pg 362, 2011). **Nature**. 2011;471.

123.   Stahlberg A, Thomsen C, Ruff D et al. Quantitative PCR Analysis of DNA, RNAs, and Proteins in the Same Single Cell. **Clinical Chemistry**. 2012;58:1682-1691.

124.   Quintana E, Shackleton M, Foster HR et al. Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. **Cancer Cell**. 2010;18:510-523.

125.   Cheung TH, Rando TA. Molecular regulation of stem cell quiescence. **Nat Rev Mol Cell Biol**. 2013;14:329-340.

126.   D'Amour KA, Agulnick AD, Eliazer S et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. **Nat Biotechnol**. 2005;23:1534-1541.

127. Norrman K, Strombeck A, Semb H et al. Distinct gene expression signatures in human embryonic stem cells differentiated towards definitive endoderm at single-cell level. **Methods**. 2012;59:59-70.
128. Shehata M, Teschendorff A, Sharp G et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. **Breast Cancer Research**. 2012;14.
129. Clark GM, Sledge GW, Jr., Osborne CK et al. Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. **J Clin Oncol**. 1987;5:55-61.
130. Liu SL, Dontu G, Mantle ID et al. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. **Cancer Research**. 2006;66:6063-6071.
131. Gupta PB, Onder TT, Jiang GZ et al. Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. **Cell**. 2009;138:645-659.
132. Al Dhaheri Y, Attoub S, Arafat K et al. Salinomycin induces apoptosis and senescence in breast cancer: Upregulation of p21, downregulation of survivin and histone H3 and H4 hyperacetylation. **Bba-Gen Subjects**. 2013;1830:3121-3135.
133. An H, Kim JY, Lee N et al. Salinomycin possesses anti-tumor activity and inhibits breast cancer stem-like cells via an apoptosis-independent pathway. **Biochem Biophys Res Commun**. 2015;466:696-703.
134. D'Amato NC, Ostrander JH, Bowie ML et al. Evidence for phenotypic plasticity in aggressive triple-negative breast cancer: human biology is recapitulated by a novel model system. **PLoS One**. 2012;7:e45684.
135. Su Y, Subedee A, Bloushtain-Qimron N et al. Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. **Cell Rep**. 2015.
136. Leder K, Holland EC, Michor F. The Therapeutic Implications of Plasticity of the Cancer Stem Cell Phenotype. **Plos One**. 2010;5.
137. Shaw FL, Harrison H, Spence K et al. A detailed mammosphere assay protocol for the quantification of breast stem cell activity. **J Mammary Gland Biol Neoplasia**. 2012;17:111-117.
138. Schachter M. Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update. **Fundamental & clinical pharmacology**. 2005;19:117-125.
139. Clendening JW, Penn LZ. Targeting tumor cell metabolism with statins. **Oncogene**. 2012;31:4967-4978.
140. Engstrom MJ, Opdahl S, Hagen AI et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. **Breast Cancer Res Tr**. 2013;140:463-473.