European School of Molecular Medicine

(SEMM) - Naples site

Università degli Studi di Napoli "Federico II"

PhD in Molecular Medicine –VI/XXIV cycle

Molecular Oncology

Identification of new markers for the characterization and isolation of breast cancer stem cells

Felicia Leccia

Tutor: Prof. Francesco Salvatore

Internal Supervisor: Prof. Luigi Del Vecchio

External Supervisor: Dr. Stéphane Ansieau

Academic Year: 2011/2012

TABLE OF CONTENTS

ABBREVIATIONS

ABCG2	ATP-Binding Cassette G2
ALCAM	Activated Leukocyte Cell Adhesion Molecule
APC	Allophycocyanin
BCRP1	Breast Cancer Resistance Protein 1
BCSC	Breast Cancer Stem Cell
BLBC	Basal-like breast cancer
CALLA	Common Acute Lymphoblastic Leukaemia Antigen
CD	Cluster of Differentiation
cDNA	Complementary DNA
CEACAM6	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6
CEACAM8	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 8
CK	Cytokeratin
CSC	Cancer Stem Cell
CXCR4	C-X-C Chemokine Receptor Type 4
DAF	Decay Accelerating Factor
DCIS	In situ Ductal Carcinoma
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DPPIV	Dipeptidyl peptidase IV
EGFR	Epidermal Growth Factor Receptor
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-mesenchymal transition
EpCAM	Epithelial Cell Adhesion Molecule
ER	Estrogen Receptor
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal bovine serum
FCM	Flow Cytometry
FITC	Fluorescein isothiocyanate
FSC	Forward Scatter

FSC-A	Forward scatter-Area
FSC-H	Forward scatter-Height
H-CAM	Homing-associated Cell Adhesion Molecule
HER2	Human Epidermal Receptor 2
HLA	Human Leukocyte Antigen
HMEC	Human Mammary Epithelial Cell
HSA	Heat Stable Antigen
HPRT1	Hypoxanthine phosphoribosyltransferase
ICAM	Intercellular Adhesion Molecule
IHC	ImmunohistochemIstry
IDC	Infiltrating or Invasive Ductal Carcinoma
IMDM	Iscove's Modified Dulbecco's Medium
Ma-CFC	Human Mammary epithelial Colony-Forming Cell
MaSC	Mammary Stem Cell
MDR	Multidrug resistance
MFE	Mammosphere Forming Efficiency
MFI	Mean Fluorescence Intensity
MGC-24	Multiglycosylated core protein-24
MIRL	Membrane Inhibitor of Reactive Lysis
MoAb	Monoclonal Antibody
mRNA	Messenger RNA
MRUs	Mammary Repopulating Units
MUC	Mucin
NOD/SCID	Nonobese Diabetic/ Severe Combined Imunodeficincy
OCT4	Octamer-binding transcription factor 4
PBS	Phosphate-buffered saline
PE	Phycoerythrin
PE-Cy7	Phycoerythrin-Cyanin 7
PerCP	Peridinin Chlorophyll Protein
PETA-3	Platelet Endothelial Tetraspan Antigen 3
PgR	Progesteron Receptor
PMPI	p-Maleimidophenyl isocyanate
RNA	Ribonucleic acid

RPMI	Roswell Park Memorial Institute medium
RT-qPCR	Real-time reverse transcription Polymerase Chain Reaction
SD	Standard Deviation
SEM	Standard Error of the Mean
shRNA	Short hairpin RNA
SMA	Smooth Muscle Actin
SP	Side Population
SSC	Side Scatter
TAPA-1	Target of Anti-Proliferative Antibody-1
TNBC	Triple Negative Breast Cancer
TPN/TN	Triple Negative

LIST OF FIGURES AND TABLES

ABSTRACT

Breast cancer is the first human carcinoma for which a putative cancer stem cell subpopulation (BCSC) has been isolated with the $CD44^{+}/CD24^{-/low}$ phenotype. However, several studies have highlighted that $CD44^{+}/CD24^{-/low}$ cannot be considered the universal marker phenotype of BCSCs. The aim of this thesis is to identify novel BCSCs markers. We focused our work on the basal-like breast cancers that seem to derive from a developmental stage of mammary epithelial cell that is different from the primitive stem cell, namely the luminal progenitor.

In the first part of the present thesis we have characterized breast cancer cell lines resembling different tumor molecular subtypes, by using polychromatic flow cytometry. These analyses led us to hypothesize that the basal-A cell line HCC1937 could contain a putative $CD24^{+}CD338^{+}$ CSC subpopulation. We have next found that CD338 overexpressing cells, namely $CD338^{bright}$, overlap, almost completely, with the side population in the HCC1937 cell line. Since side population is a property of stem/progenitor cells, this result suggested a key role of CD338 in determining stem-like properties of HCC1937 cell line.

To test this hypothesis, we next isolated the putative BCSCs, on the base of CD24 and CD338 expression, and we further studied them to explore their stem-like and tumorigenic properties. We found that $CD24^{hi}$ cells display a higher mammosphere forming efficiency than $CD24^{-}$ cells and, among $CD24^{hi}$ cells, CD338 overexpressing cells are able to form mammospheres with higher efficiency than $CD338^{-}$ cells. This result confirmed a relevant role of CD338 in stem-like properties of HCC1937 cell line. Furthermore, the soft-agar colony assays revealed that CD338 positive cells are transformed in contrary to the negative ones. We next evaluated the tumorigenic potential of the different CD338 sorted subpopulations and we found that cells expressing CD338 at an intermediate level, namely $CD338^{low}$, were the most

tumorigenic ones. Furthermore, the analysis of $CD338^{bright}$ and $CD338^{low}$ sorted sub-populations, after their growth in culture for several weeks, revealed that $CD338^{bright}$ population is able to divide asymmetrically giving rise to its $CD338^{low}$ progeny which, in contrast, is not able to divide asymmetrically. This result suggested that $CD338^{bright}$ population could constitute the most immature population and its $CD338^{low}$ progeny the more differentiated progenitor cells. Taken together, these results strongly support the presence in the HCC1937 cell line of a putative BCSC subpopulation overexpressing CD338 and bearing the phenotype $CD44^{+}CD24^{+}CD338^{bright}$, i.e. an antigenic combination different from the classical $CD44^{+}/CD24^{-/low}$.

Concurrently to the study of breast cancer cell lines, a complementary work was carried out to standardize the flow-cytomertic methods to study primary cultures established from human breast cancer tissues. This work will be subsequently used to isolate and study the $CD44^{+}CD24^{+}CD338^{bright}$ CSC population from human breast cancer tissues. In this study, we have also tried to determine whether patient-derived breast cancer cell cultures maintain the cellular heterogeneity of primary tissues and may therefore be used for *in vitro* modeling of breast cancer subtypes. To this aim, we used a much larger vocabulary of surface markers than those used in previous studies to characterize primary cultures. Most of surface antigens analyzed were heterogeneously expressed. On the other hand, breast cancer cell cultures displayed concomitant high expression of the basal marker CD10/CALLA and low expression of CD326/EpCAM. Furthermore, we found that they express CK5, SMA and vimentin, and are weakly positive for CK19 and CK18-negative. Our results, while confirming that *in vitro* culturing of breast cancer cells reduces luminal lineage-type of cells, indicate the increased propensity for the selection of myoepithelial/basal breast cancer cells. Furthermore, we have demonstrated that breast cancer cell cultures preserve inter-tumor heterogeneity and express stem/progenitor markers that can be identified, quantified and categorized by flow cytometry.

I. INTRODUCTION

1.1 The Cancer Stem Cell theory

1.1.1 Two general models of heterogeneity in solid cancers

It is well established that tumors are proliferatively heterogeneous. Particularly, cancer cells within a single tumor exist in multiple states of differentiation and only a subfraction of tumor cells is able to self-renew either *in vitro* or in *in vivo* xenograft models, by generating tumors with serial transplantations. Two theories have been proposed to explain this (Figure 1.1a). The "stochastic" theory suggests that cancers arise from waves of genetic and epigenetic mutations that affect the survival and reproduction of cells, providing a selective advantage and generating clonal heterogeneity. In contrast, the "cancer stem cell" (CSC) theory holds that tumors are organized in a cellular hierarchy in which cancer stem cells (CSCs) are the only cells with unlimited proliferation potential and with the capability of driving tumor growth and progression. Thus, CSCs can be defined as rare cells within tumors with the ability to self-renew and give rise to the phenotypically diverse cancer cell population.

In contrast to the "stochastic" model of oncogenesis, where transformation results from random mutations and subsequent clonal selections, according to the CSC model cancers would originate from the malignant transformation of a stem or progenitor cells through the deregulation of the normally tightly regulated self-renewal program or from transformation of committed cells through dedifferentiation of mature cells that gain a self-renewal potential. So, the origin of CSCs is currently elusive and, as a consequence, the term "cancer stem cell" is disputed because it is not clear weather normal stem cells are the natural precursors of "cancer stem cells". The origin of CSCs seems to be tissue dependent. As regard breast cancer, recent studies have shown that in transformed mammary epithelial cells, induction of epithelial-mesenchymal transition (EMT) that is a transition of epithelial cells to a less

differentiated stage, creates populations of cells that are highly enriched for CSCs properties, such as tumor-seeding ability, mammosphere formation and stem cell surface marker expression (Mani *et al*., 2008; Morel *et al*., 2008).

1.1.2 Normal stem cells and cancer stem cells

Normal stem cells (NSCs) and cancer stem cells share a number of important characteristics. Particularly, they both have the capacity of self-renewal and extensive proliferation. In the case of tumor cells, this takes the form of self-sufficiency in growth signaling and uncontrolled cellular proliferation, whereas for normal stem cells, this is a tightly controlled process that occurs during embryogenesis, organogenesis, and maintenance and repair of adult tissues. NSCs and CSCs also share the ability to differentiate, giving rise to their progenity in tissue and to tumor heterogeneity, respectively. Moreover, both cell types have a resistance to environmental toxins, chemotherapeutic and radiation agents, often as a result of multidrug resistance (MDR) via expression of ABC (ATP-Binding Cassette) family of transporter proteins. Stem cells and tumor cells also share the characteristic of being able to survive in anchorage independent conditions (resistence to anoikis), leading to migration and homing for stem cells and potentially to metastatic disease for tumor and cancer stem cells. Finally, both cell types are long-lived with active anti-apoptotic pathways and telomerase activity. The longevity of normal stem cells increases the risk of these cells, over a lifetime, to acquire the multiple genetic mutations necessary for tumorigenesis. This would lead to consider normal stem cells as appropriate candidates for tumor initiating but, on the other hand, the slow-cycling property of stem cells is not in agreement with this consideration, since DNA replication is likely a prerequisite for mutation.

1.1.3 Implication of the CSC theory in cancer treatment

The CSC theory has implications in cancer treatment. Current anticancer therapy is effective at debulking the tumor mass but treatment effects are transient, with often occurring tumor relapse and metastatic disease. A possible explication for the failure of existing therapies is that they fail to kill cancer stem cells (Figure 1.1b). As mentioned before, normal stem cells from various tissues tend to be more resistant to chemotherapeutics than mature cell types from the same tissues (Harrison and Lerner 1991). One of the reasons that could explain this, is that stem cells express high levels of anti-apoptotic proteins (Feuerhake *et al*., 2000) and ABC transporter proteins (Zhou *et al*., 2001). Also cancer stem cells seem to overexpress transporter proteins and this could explain that these cells would be more resistant to chemotherapeutics than committed/differentiated tumor cells with limited proliferative potential. Even therapies that cause complete regression of tumors might spare enough cancer stem cells to allow regrowth of the tumors. Targeting differentiated as well as quiescent tumor stem cells is a prerequisite for therapy to be efficient. On the other hand, it is noteworthy to consider the recent emerging possibility of the existence of an equilibrium between CSCs and non-CSCs within tumors. In particular, whereas CSCs can differentiate into non-CSCs giving rise to the tumor heterogeneity, the reverse process would be possible: non-CSCs, could be reprogrammed into CSCs. This phenotypic plasticity has implications for cancer treatment: if non-CSCs can give rise to CSCs, therapeutic elimination of CSCs may be followed by their regeneration from residual non-CSCs, allowing tumor regrowth and clinical relapse. Therefore, optimal therapeutic approaches would target both CSCs and differentiated cells (Gupta *et al*., 2009).

1.1.4 Cancer stem cells in solid cancers

The existence of cancer stem cells was first proven in the context of acute myeloid leukemia (AML). In this case, surface markers were used to distinguish AML stem cells ($CD34^+CD38^-$) from the remaining AML cells which had limited proliferative potential (Lapidot *et al*., 1994; Bonnet and Dick, 1997). These cells were found to be the only cells capable of transferring AML from human patients to NOD/SCID mice, providing evidence that not all AML cells have *in vivo* clonogenic capacity and only the small subset of CSCs was capable of regenerating the cancer. Many groups have extrapolated the CSC hypothesis from the haematopoietic system to solid cancers and, although the evidence for CSCs in solid cancers is in its infancy compared to the haematopoietic field, the body of supporting data is growing rapidly. The first solid CSCs were identified in breast tumors (Al-Hajj *et al.,* 2003), and then CSCs were isolated from brain (Singh *et al*., 2003), colon (O'Brien *et al*., 2007), melanoma (Fang *et al*., 2005), pancreatic (Hermann *et al*., 2007), prostate (Collins *et al*., 2005), ovarian (Bapat *et al*., 2005), lung (Ho *et al*., 2007) and gastric (Fukuda *et al*., 2009) cancers. Cells with CSC characteristics from human brain tumors (glioblastomas) were first isolated using a clonogenic sphere culture technique to produce the so called neurospheres (NS) (Singh *et al*., 2003; Singh *et al*., 2004). These NS cells are highly enriched for cell surface marker CD133 and nestin (a neural stem cell marker), have a marked capacity for proliferation and self-renewal and are capable of *in vitro* differentiation into phenotypes identical to the tumor *in situ*. Breast cancer is the first human carcinoma for which a putative cancer stem cell subpopulation (BCSC, breast cancer stem cell) has been isolated by Clarke and colleagues. This study showed that in nine breast cancer samples, a minority of cells, bearing the surface marker phenotype $CD44^+CD24^-$, were capable of generating tumors in NOD/SCID mice even when implanted in low numbers. By contrast, the other cancer cell populations, such as $CD44^+CD24^+$, failed to generate tumors even when implanted in high numbers (Al-Hajj *et al.,*

2003). Cell surface antigens have been used with good success in the haematological system to identify stem and progenitor cells, but in solid tumors, such as breast cancers, there is still the lack of reliable markers. The relevance of CD44 and CD24 in identifying breast CSCs is still matter of debate and, actually, $CD44^{+}CD24^{-/low}$ not considered as a universal antigenic phenotype for isolation of BCSCs. Furthermore, the use of surface antigens to classify different cell subpopulations is actually highly disputed. The major drawback of this technique is that there is rarely a marker that is 100% expressed in one population and completely lost in the next. Indeed a gradient of expression of cell surface antigens is more often detected. As a consequence, cell sorting based on surface antigens allows for enrichment rather than isolation of the various sub-populations.

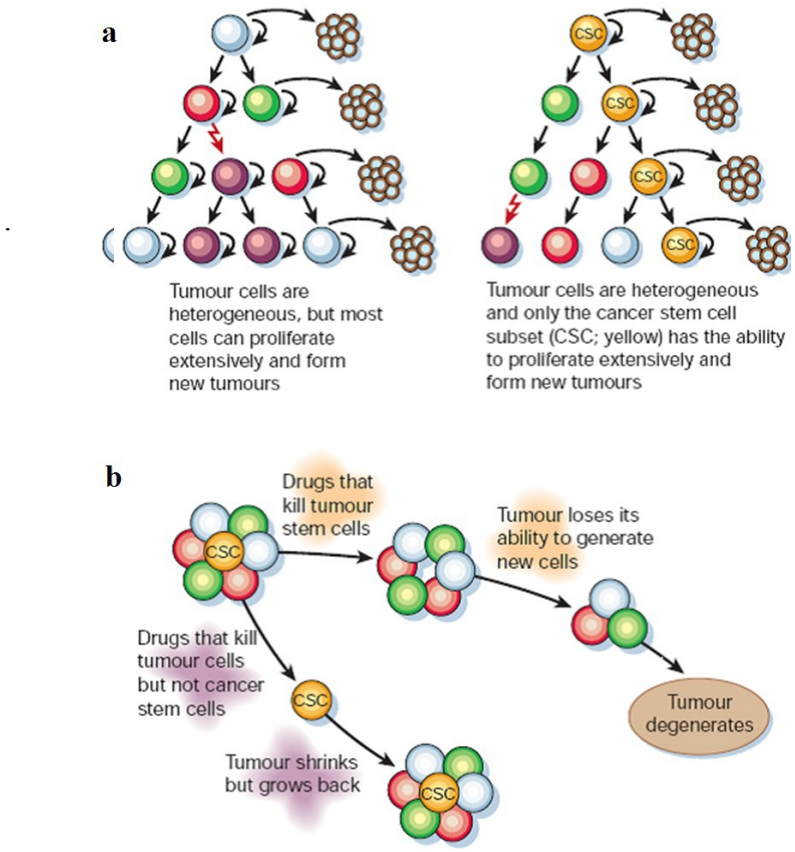


Figure 1.1. The cancer stem cell theory. (a) Two general models of hete8rogeneity in solid cancers; **(b)** Comparison between conventional therapies and cancer stem cell targeted therapies (Reya, 2001).

1.2 The human mammary gland and stem cells in normal human breast

The mammary gland is a compound tubulo-alveolar composed of a series of branched ducts that drain sac-like alveoli (lobules) during lactation. In the mouse, the mammary epithelium is embedded within a mammary fat pad, whereas in humans, it is embedded within a fibrous and fatty connective tissue. The functional units of the human mammary gland are comprised of terminal ducts and alveoli, which together form the terminal duct lobular units (TDLUs). Collectively, TDLUs form the branches of a greater ductal-lobular system composed of an inner layer of polarized luminal cells and an outer layer of myoepithelial cells (Figure 1.2). Myoepithelial cells are contractile cells that form a sheath around the ductal network of the breast and are characterized by expression of common acute lymphoblastic leukaemia antigen (CALLA) also known as CD10 (Gusterson *et al*., 1986), Thy-1 also known as CD90 (Gudjonsson *et al*., 2002), alpha-smooth muscle actin (α-SMA) (Gugliotta *et al.*, 1988), vimentin (Guelstein *et al*., 1988), and cytokeratin (CK) 5 and CK14 (Taylor-Papadimitriou *et al*., 1989). Luminal epithelial cells are cuboidal/columnar cells that line the ducts and alveoli and are the cells responsible for milk production during lactation. They are characterized by expression of Mucin-1 (MUC-1) also known as CD227 (Petersen *et al*. 1986), epithelial cell adhesion molecule (EpCAM) also known as epithelial surface antigen (ESA) or CD326 (Latza *et al*., 1990), and CK7, CK8, CK18, and CK19 as well as estrogen receptor (ER) and progesterone receptor (PgR). The adult human breast is composed of 15 to 20 lobes each with multiple lobules surrounded by adipose tissue. Additionally, the breast has a system of lymphatic vessels responsible for draining breast tissue leading to internal mammary lymph nodes and axillary regional lymph nodes.

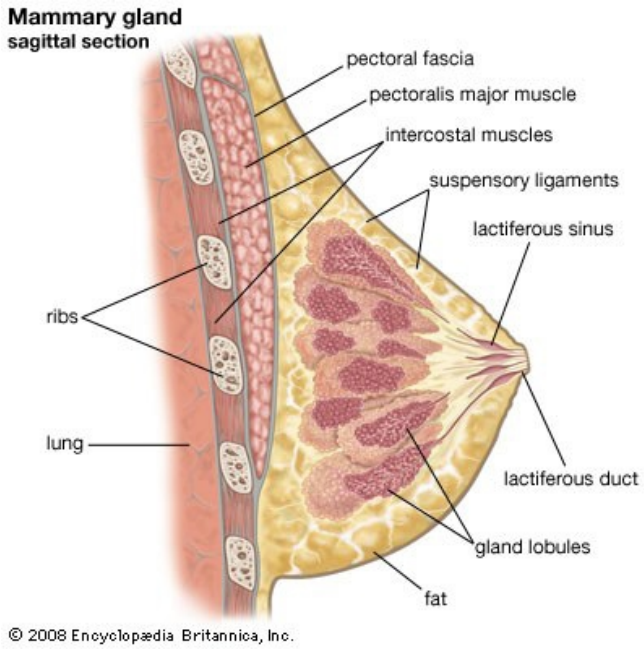


Figure 1.2. Schematic representation of the human mammary gland.

The human breast is a dynamic gland with tissue homeostasis occurring during early development, puberty, within menstrual cycles, during pregnancy and lactation, and eventual involution during menopause. During periods of pregnancy and lactation, the breast goes through further rounds of development with an increase in cell growth and formation from the luminal epithelial lineage of functional milk-secreting alveoli. Following these periods and again during menopause, there is an involution through apoptosis of the breast tissue. Generation and maintenance of the mammary epithelium is hypothesized to be via the mammary stem cell (MaSC), which is defined as that cell which can generate both the ductal and lobular structures of the mammary gland. (Kordon and Smith, 1998; Stingl *et al*., 1998). Transplantation of a single FACS-isolated MaSC, can reconstitute, a normal mammary gland (Stingl *et al*., 2006: Shackleton *et al*., 2006). Transplantation studies have an important limit: they mimic a regenerative state that in certain circumstances forces stem cells to differentiate

into lineages for which they usually do not contribute to under physiological conditions. So, it is necessary to develop assays that allow to define the cellular hierarchy of the mammary gland during physiological conditions. To this aim, genetic lineage-tracing approaches in mice have been recently developed. In this study, it has been demonstrated that the mammary gland contains different types of long-lived stem cells. Particularly, Blanpain and colleagues have found that the expansion and maintenance of luminal and myoepithelial lineages is ensured by the presence of two types of lineage-restricted stem cells, rather than by rare multipotent stem cells (Van Keymeulen *et al*., 2011). The demonstration that the mammary gland contains different types of long-lived stem cells, has important implications for the study of the origin of breast cancers.

The study of stem cells within the human mammary epithelium has been much more difficult than their mouse counterparts due to the lack of a suitable *in vivo* xenotransplantation assay. The human mammary epithelium *in viv*o is embedded within an irregular extracellular matrix rich in collagen fibers and mammary fibroblasts. In rodents, the extracellular matrix is composed primarily of adipocytes. Consequently, human mammary epithelial cells (HMECs) fail to recapitulate the ductal-lobular outgrowths when transplanted into cleared fat pads of immune-deficient mice, presumably as a result of inappropriate epithelial-stromal interactions (Hovey *et al*., 1999). A solution to circumvent this, is to transplant single cell suspensions of HMECs into humanized fat pads (i.e. fat pads pre-inoculated with human mammary fibroblasts) or have embedded the HMECs within collagen gels and/or Matrigel and place these cells either subcutaneously or under the renal capsule of immune-deficient mice (Yang *et al*., 1994; Stingl *et al*., 2003; Proia *et al*., 2006). HMECs transplanted under such conditions recapitulate histologically normal looking human mammary epithelium, complete with polarized luminal cells, SMA-expressing myoepithelial cells, basal clear cells and small light cells (Stingl *et al*., 2003). The most useful markers for isolating human mammary stem

cells are the epithelial cell adhesion molecule (EpCAM/CD326), the α6-integrin/CD49f and, to a lesser degree, the luminal cell-specific glycoprotein MUC1/CD227, the common acute lymphoblastic leukemia antigen CALLA/CD10 and the GPI-linked glycoprotein Thy-1/CD90 (Stingl *et al*., 1998; Stingl *et al.,* 2001; Eirew *et al*., 2008; Villadsen *et al*., 2007; Raouf *et al*., 2008). EpCAM/CD326 is an epithelial-specific molecule that is expressed at high level by luminal cells and at lower level by basal cells; CD49f is an integrin known to participate in cell adhesion as well as cell surface mediating signaling and it displays a pattern of expression inverse to that of EpCAM (Stingl *et al*., 1998; Carter *et al*., 1990). Although both EpCAM/CD326 and CD49f are not particularly useful for identifying different subsets of HMECs when separately used, when combined they permit a number of distinct cell types to be resolved by flow cytometry. MUC1/CD227 is a cell surface glycoprotein expressed by most epithelial cells and involved in cell-cell and cell-substrate adhesion. CALLA/CD10 is a membrane-associated endopeptidase overexpressed in many neoplastic cell types and Thy1/CD90 is a membrane-associated glycoprotein involved in cell adhesion, migration and proliferation. *In vitro* analyses of human mammary epithelial colony-forming cells (Ma-CFCs) have demonstrated the existence of three distinct progenitors within the human mammary epithelium (Figure 1.3): 1) the luminal-restricted progenitors which generate colonies that are composed of cells with a luminal cell phenotype (expressing CK8, 18, 19 and 9). Replating of these colonies in secondary CFC assays reveals that cells of these colonies are indeed restricted to a luminal cell fate, since they only form pure luminal cell colonies. These luminal restricted progenitors display the phenotype EpCAM/CD326^{hi} CD49f^{+} MUC1/CD227^{+} CALLA/CD10^{-} Thy1/CD90^{-} (Stingl *et al*., 2001; Villadsen *et al*., 2007). 2) The bipotent progenitors display a EpCAM/CD326$^{-/low}$ CD49f^{high} MUC1/CD227^{-} CALLA/CD10^{+} Thy1/CD90^{+} antigenic phenotype. These progenitors generate colonies that are characterized by a central core of CK14−CK18+CK19+MUC1+ luminal cells and a

peripheral zone of CK14+ basal cells. Single-cell cloning of these progenitors confirms that these mixed colonies are clonal in origin (Stingl *et al*., 2001). 3) The myoepithelial-restricted progenitors generate colonies that are composed solely of CK14+CK18−CK19−MUC1− basal-like cells. Serial passaging of the colonies generated by bipotent progenitors has demonstrated that myoepithelial-restricted progenitors arise from bipotent progenitors. The unipotent progenitors, also known as MRUs (Mammary Repopulating Units), have the same antigenic phenotype as bipotent progenitors, namely EpCAM/CD326$^{-/low}$ CD49f^{high} MUC1/CD227^{-} CALLA/CD10^{+} Thy1/CD90^{+}, suggesting a basal position within the mammary epithelium. However, it is not known whether bipotent progenitors and MRUs types represent distinct or overlapping populations within the basal cell population (Stingl *et al.,* 2001; Eirew *et al*., 2008; Raouf *et al*., 2008).

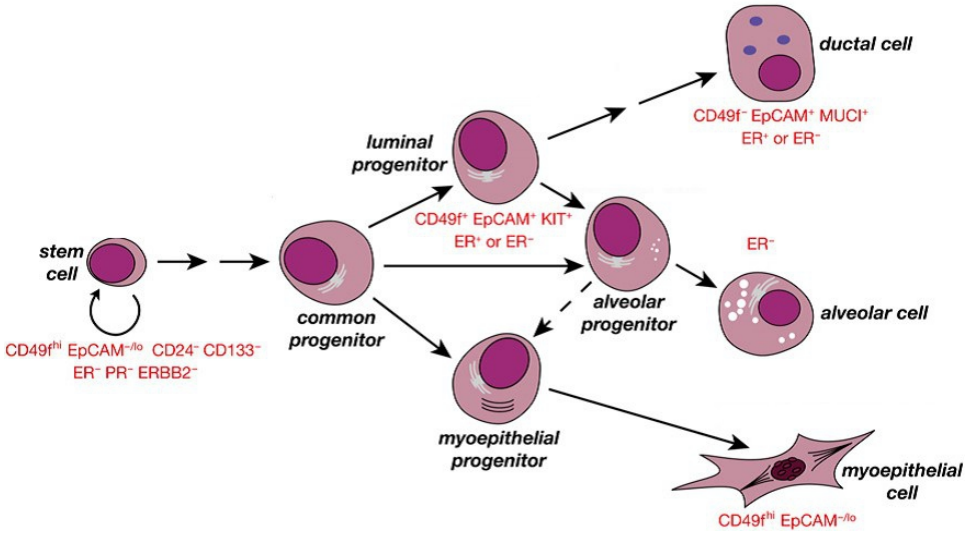


Figure 1.3. Model of the differentiation hierarchy within human mammary epithelium. The common progenitor is also referred to as a bipotent progenitor cell (Visvader, 2009).

1.3 Breast cancer: a very heterogeneous disease

Breast cancer is a genetically and clinically heterogeneous disease. Many breast cancer classification systems have been developed over many decades and have evolved into a tool that is used to aid in treatment and prognosis. The different breast cancer classification systems are based on several tumor-intrinsic features that separate breast tumors into multiple subtypes, including histological and immunopathological subtypes, molecular subtypes and the more recently described functional subtypes based on the identification and characterization of tumor initiating cancer stem cells in the different molecular subtypes of breast cancer.

1.3.1 Hystological and immunopathological classification of breast cancers

Based on their overall morphology and structural organization, breast cancers can be categorized into *in situ* carcinoma and invasive or infiltrating carcinoma (Figure 1.4a). *In situ* breast carcinoma is further sub-classified as either ductal or lobular, growth patterns and cytological features form the basis to distinguish between the two types. *In situ* ductal carcinoma (DCIS) is considerably more common than *in situ* lobular carcinoma (LCIS) and includes a heterogeneous group of tumors. Particularly, DCIS has traditionally been further sub-classified on the basis of the architectural features of the tumor which has given rise to five well recognized subtypes: Comedo, Cribiform, Micropapillary, Papillary and Solid (Connolly *et al*., 2004; Figure 1.4a). Similar to *in situ* carcinomas, invasive carcinomas are a heterogeneous group of tumors differentiated into histological subtypes. The major invasive tumor types include infiltrating ductal carcinoma (IDC) (Figure 1.4b), invasive lobular (Figure 1.4c), ductal/lobular, mucinous (colloid), tubular, medullary and papillary carcinomas. Among these tumor types, IDC is, by far, the most common subtype accounting for 70–80% of all invasive lesions (Li *et al*., 2005). IDC is further sub-classified as either

well-differentiated (grade 1), moderately differentiated (grade 2) or poorly differentiated (grade 3) based on the levels of nuclear pleomorphism, glandular/tubule formation and mitotic index (Figure 1.4a). While this classification scheme has been a valuable tool for several decades, it relies solely on histology without using newer molecular markers that have a proven prognostic significance.

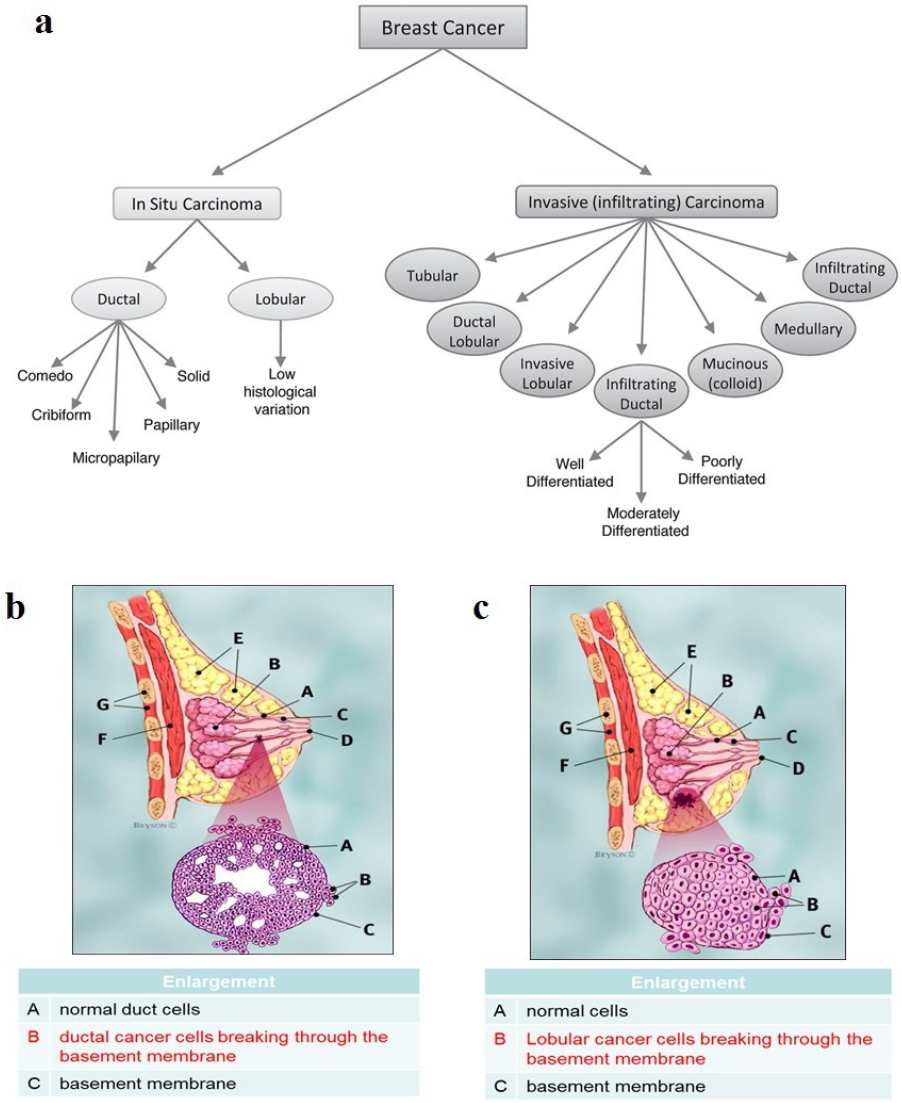


Figure 1.4. Hystological subtypes of breast cancer. (a) Hystolagical classification of breast cancer (Malhotra *et al*., 2010). **(b,c)** Schematic representation of invasive ductal carcinoma (IDC, b) and invasive lobular carcinoma (ILC, c).

In light of surgical advances leading to breast-conserving therapy, it has become necessary to more accurately stratify patients based on relative risk of recurrence or progression. These demands have led to the generation of several newer classification systems that incorporate specific markers such as estrogen receptor (ER), progesterone receptor (PR or PgR) and human epidermal receptor 2 (HER2/ErbB2). In contrast to DCIS, where the use of molecular markers is still debated, the utility of ER, PR and HER2/neu is well accepted for IDC and it is recommended that their status be determined on all invasive carcinomas. The status of these markers helps in determining which patients are likely to respond to targeted therapies (Malhotra *et al*., 2010). In particular, the ER status is used to identify tumors that may respond to anti-estrogen (endocrine) therapy, including ER antagonists or aromatase inhibitors, which target ER-dependent signaling (Jordan *et al*., 2007). The PR status, generally correlated with the ER status, has less clinical significance. However, the PR status does not appear to predict relative benefit from specific types of endocrine therapy (Bartlett *et al*., 2011), and overall, ER+/PR+ cases may not receive additional benefit from endocrine therapy compared with ER+/PR– cases (Stendhal *et al*., 2006). HER2+ cases are treated with targeted therapies such as the monoclonal antibody trastuzumab, which binds to HER2, mediates antibody-dependent cytotoxicity, and disrupts HER2-dependent signaling (Junttila *et al*., 2009). There is currently no standard targeted therapy for cases assessed as ER– and HER– by IHC (immunohistochemistry), although this represents an intensive area of research. Thus, combinations of these markers allow for the assignment of individual cases to specific categories, namely ER+ (ER+/HER2–), HER2+ (ER–/HER2+), triple negative (TN or TPN; ER–/PR–/HER2–), and triple positive (ER+/PR+/HER2+). From a prognostic point of view, ER+ tumors exhibit the best overall outcome, whereas HER2+ tumors, previously associated with poor outcome, now exhibit an improved overall outcome, after the advent of HER2-

targeted therapy (Smith *et al*., 2007). TN tumors, on the other hand, are linked to the worst prognosis among the subtypes (Nishimura *et al*., 2008).

1.3.2 Molecular breast cancer subtypes

Over the past decade, the advent of high-throughput/high-content microarray-based gene expression profiling technologies has facilitated large-scale studies of breast cancer cohorts, leading to the identification of multiple molecular subtypes of breast cancers (Figure 1.5). Studies by Perou and colleagues described five molecular subtypes defined by distinct transcriptional signatures: basal-like, ErbB2+, normal breast like, luminal subtype A and luminal subtype B (Sorlie *et al*., 2003; Perou *et al*., 2000; Sorlie *et al*., 2001). More recently, a new subtype classified as "claudin low" has also been identified (Herschkowitz *et al*., 2007; Prat *et al*., 2010). Molecular subtypes defined by Perou and colleagues partially recapitulate the immunopathological classes, while adding an additional level of detail. Two luminal subtypes (A and B) contain principally ER+ cases and are distinguished by the presence of genes regulated by the ER signaling pathway. The luminal A subtype is associated with higher levels of *ESR1* (estrogen receptor gene 1), ER, and ER-regulated genes, decreased proliferation and improved overall outcome (Sorlie *et al*., 2001; Perou *et al*., 1999; Perreard *et al*., 2006). Luminal B tumors appear to exhibit decreased levels of *ESR1*, ER, and ER-regulated genes as well as increased proliferation and relatively worse prognosis. Thus, this molecular classification was able to stratify the ER+ population into several subtypes that demonstrated a difference in patient survival. This is significant because, even though clinical assessment of IDC utilizes ER, PR and ErbB2 status, these markers did not allow separation of the two distinct ER+ subtypes (i.e., Luminal A and Luminal B) that have very different clinical outcomes. Three molecular subtypes contain predominantly ER– cases (Perou *et al*., 2000): 1) the molecular ErbB2+ subtype generally overlaps with IHC-defined HER2+ tumors

(ER-/HER2-) (Perou *et al*., 2000); the basal-like subtype broadly corresponds to the TN (ER–/PR–/HER2–) cohort (Nielsen *et al*., 2004); 3) the claudin-low subtype linked to metaplastic breast cancers and high risk of recurrence, which shows similarity to stem cell-linked and epithelial to mesenchymal transition-linked (EMT-linked) signatures (Hennessy *et al*., 2009). Finally, the normal-like subtype, as its name implies, displays similar expression pattern to normal breast tissue. The significance of this subtype has yet to be determined, and some argue that it may represent a mere contamination of samples with normal breast tissue (Parvin *et al*., 2012; Huang *et al*., 2012; Parker *et al*., 2009).

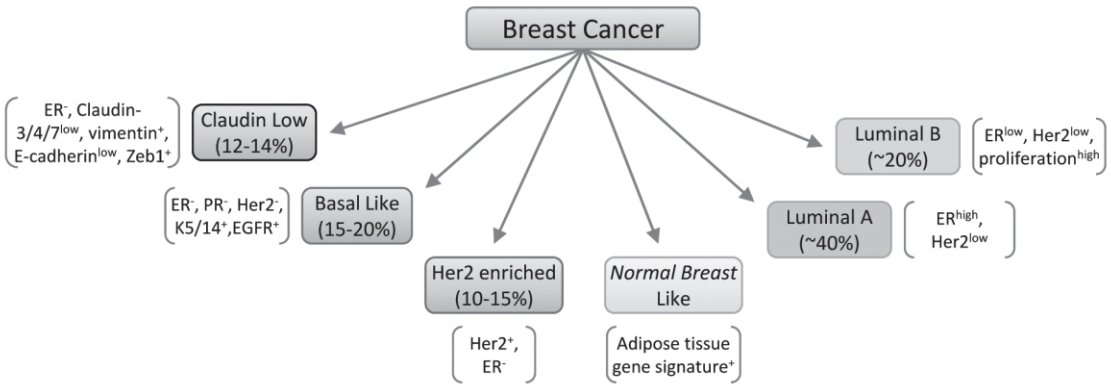


Figure 1.5. Molecular classification of breast cancers (Malhotra *et al.*, 2010).

1.3.3 Triple negative, basal-like and BRCA1-mutated breast cancers

As mentioned before, the molecular subtypes of breast cancer display highly significant differences in prediction of overall survival with the basal like/triple-negative (ER-/PR-/ErbB2-) subtype having the shortest survival. Basal-like breast cancers (BLBCs) represent 10%-25% of all breast tumors, depending on the demographics of the population. These tumors were so named because their neoplastic cells consistently express genes usually found in normal basal/myoepithelial cells of the breast including high molecular weight basal cytokeratins (CK; CK5/6, CK14 and CK17), vimentin and P-cadherin (Nielsen *et al*., 2004; van de Rijn *et al*., 2002; Gusterson *et al*., 2005; Lakhani *et al*., 2001). BLBCs often affect younger patients, frequently lack expression of hormone receptors and HER2, show *TP53* gene mutations in up to 85% of cases, display exceedingly high level of proliferation related genes and express epidermal growth factor receptor (EGFR) in >60% of cases (Perou *et al*., 2000; Sorlie *et al*., 2001; Hu *et al*., 2006). The high proliferative potential of BLBCs is largely associated with loss of function of RB1 and p53, critical regulators of cell cycle. Basal like cancers are also frequently harboring *BRCA1* mutation. It is believed that this combined alteration of p53/RB and BRCA1 is also responsible for the high euploidy observed in these tumors, including a huge number of chromosomal changes and translocations (Perou, 2011). Despite the great interest in basal-like cancers, there is still no internationally accepted definition of these tumors. Actually, microarray based expression profiling analysis is the most used tool for the identification of basal-like breast cancers (Rakha *et al*., 2007). However, this technology is unlikely to be introduced among diagnostic tools for breast cancer patients, and results of microarray-based expression profiling using RNA extracted from formalin-fixed archival samples are not optimal (Penland *et al*., 2007). Therefore, several attempts to define an immunohistochemical surrogate for basal-like cancers have been described. The best example to date is the panel proposed by Nielsen and colleagues, where

basal-like cancers are defined as those lacking both ER and HER2 expression and expressing CK5/6 and EGFR. This panel has a specificity of 100% and a sensitivity of 76% for the identification of basal-like cancers (Nielsen *et al*., 2004).

Triple-negative breast cancers (TNBCs) account for 10–17% of all breast carcinomas (Harris *et al*., 2006; Haffty *et al*., 2006; Dent *et al*., 2007) depending on the thresholds used to define ER and PR positivity and the methods for HER2 assessment. The majority of triple-negative cancers are high-grade invasive ductal carcinomas of no special type, metaplastic carcinomas and medullary cancers. The main characteristics of triple negative cancers that have emerged from the literature illustrate the similarities between basal-like and triple-negative tumors including the fact that they more frequently affect younger patients (<50 years) (Haffty *et al*., 2006; Dent *et al*., 2007), are more prevalent in African-American women (Morris *et al*., 2007) and are significantly more aggressive than tumors pertaining to other molecular subgroups (Haffty *et al*., 2006; Dent *et al*., 2007). Although there is a great deal of overlap between triple-negative and basal-like breast cancers, this overlap is not complete. Most, but not all, BLBCs are TN and *vice versa* (Figure 1.6). Particularly, analysis of ER, PR and HER2 status of breast cancers classified by microarray-based expression profiling analysis, as pertaining to the basal-like subgroup, has revealed that 15-54% of them express at least one of these markers. On the other hand, only 56-84% of triple negative tumors express basal markers such as basal CKs and EGFR. Thus, triple-negative and basal-like breast cancers are not synonyms. However, patients with triple-negative breast cancer expressing a basal phenotype have a significantly shorter disease-free survival than those with triple-negative phenotype with a basal-like profile.

Mutation of the *BRCA1* gene confers a lifetime risk of up to 80% of developing breast cancer (Narod and Foulkes, 2004). BRCA1 is a multifunctional protein with roles in DNA damage repair, cell cycle control, transcriptional regulation and ubiquitination. It participates in

transcriptional regulation at several levels, interacting with sequence specific transcription factors, RNA polymerase II and enzymes involved in chromatin remodeling (Bochar *et al*., 2000; Zhang *et al*., 1998; Wang *et al*., 1998). BRCA1 transcriptional targets also possess tumor suppressor functions with roles in stress responses, cell cycle checkpoints, apoptosis and DNA repair (Mullan *et al*., 2005). The diverse functions of BRCA1 and its ability to suppress tumors through multiple mechanisms are highlighted by the range of mutations in *BRCA1* resulting in defects in specific and distinct functions of BRCA1, all of which result in the development of breast cancers. More recently, a role for BRCA1 in stem cell regulation and the control of mammary gland differentiation has been suggested (Foulkes, 2004). It has been proposed that a primitive stem or early progenitor cell may be the target cell for transformation and therefore represents the origin of BRCA1 related breast cancers (Liu *et al*., 2008; Lim *et al*., 2009; Molyneux *et al*., 2010; Proia *et al*., 2011).

There is increasingly evidence to suggest a link between the BRCA1 pathway and triple-negative/basal-like breast cancers (Turner *et al*., 2004; Turner and Reis-Filho 2006; Figure 1.6). In fact, the vast majority of tumors arising in *BRCA1* germ-line mutation carriers, in particular those diagnosed before 50 years of age, have morphological features similar to those described in basal-like cancers and they display a triple negative and basal-like phenotype as defined by immunohistochemistry or expression arrays (Foulkes *et al*., 2004; Lakhani *et al*., 2005; Sorlie *et al*., 2003). In particular, BRCA1-mutated breast cancers closely resemble i) basal-like breast cancer subtypes, expressing normal basal cytokeratins ii) sporadic triple negative breast cancers, lacking estrogen receptor and progesterone receptor expression and expressing low levels of Her2. Although sporadic basal-like cancers and tumors arising in *BRCA1* mutation carriers show similar molecular genetic profiles (Wessels *et al*., 2002), they differ in the lack of *BRCA1* somatic mutations in sporadic basal-like cancers. Despite this lack of *BRCA1* mutations, it has recently been demonstrated that the

BRCA1 pathway may be dysfunctional in sporadic basal-like tumors. Therefore, BRCA1 function can be abrogated both by mutation and downregulation of expression (referred to as 'BRCAness'). It has been hypothesized that this down-regulation would be mediated by epigenetic mechanisms, such as gene promoter methylation and/or transcriptional silencing of *BRCA1* (Turner *et al*., 2004; Turner and Reis-Filho, 2006).

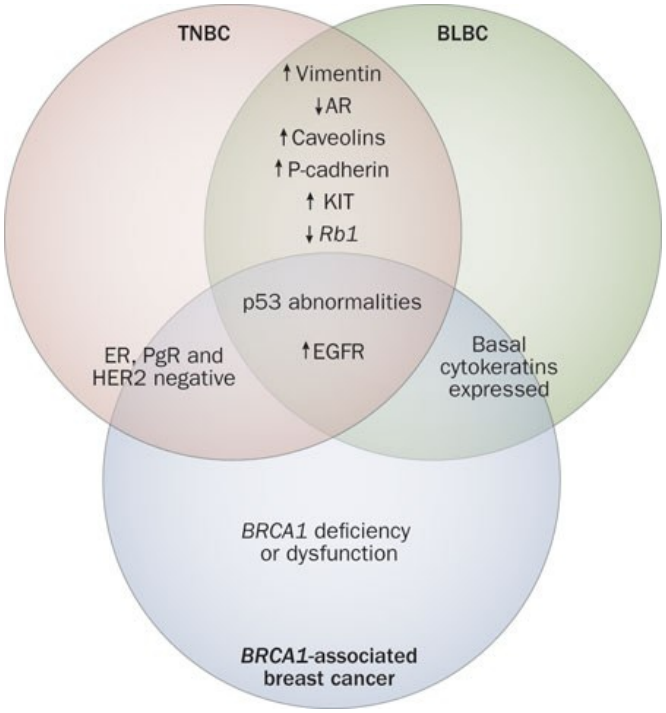


Figure 1.6. Schematic representation of the overlap between triple negative (TNBC), basal-like (BLBC) and BRCA1-muated breast cancers (Carey *et al*. 2010).

1.4 The cell of origin of breast bancers

1.4.1 Breast cancer stem cells

Al Hajj and colleagues were the first to identify a subpopulation of human breast cancer stem cells which initiated tumors in immune-deficient NOD/SCID mice (Al Hajj *et al*., 2003). They used a set of cell surface markers to sort cells with an increased tumourogenic capacity. In particular, cells which were $CD44^{+}CD24^{low}EpCAM^{+}$ and lineage- (cells lacking markers CD2, CD3, CD10, CD16, CD18, CD31,CD64 and CD140b), isolated from one primary breast cancer and eight metastases, were able to form heterogeneous tumors eight out of nine times. As few as 200 $CD44^{+}CD24^{low}EpCAM^{+}lin^{-}$ cells transplanted into NOD/SCID mice could form tumors with 100% efficiency, while $CD44^{-}CD24^{+}EpCAM^{-}$ cells formed no tumors. CD24, also known as heat stable antigen (HAS), is an adhesion molecule specifically expressed in luminal epithelial cells in human mammary gland. Thus, the low expression of luminal-specific protein CD24 in cells able to form tumors suggests a basal cell origin of tumor stem cells. The tumors contained not only the $CD44^{+}CD24^{low}EpCAM^{+}lin^{-}$ tumor-initiating cells but also the phenotypically diverse non-tumorigenic cells that comprise the bulk of tumors. A subsequent study by Ponti and colleagues carried out on 16 breast lesions (13 primary invasive carcinomas, one recurrent carcinoma and two fibroadenomas), using the sphere culture technique, resulted in the production of three long-term primary cultures which had self-renewing capacity and could differentiate into the different breast lineages (Ponti *et al*., 2005). Almost all sphere-derived cells were found to be $CD44^{+}CD24^{low}$, however, cells with self-renewal capacity only accounted for 10–20% of the total cell number, showing that only a sub-group within the $CD44^{+}CD24^{low}$ cells had self-renewal capacity. Thus, both the mammosphere culture system and the cell surface marker selection enriched for tumor initiating cells. However, the enriched rather than pure CSC population that these methods produce, and recent data suggesting that the regulation of CD24 is dynamic both *in vitro* and

in vivo (Meyer *et al*., 2009), have highlighted the need for additional markers to further discriminate the BCSC (see section 1.5c).

1.4.2 Functional classification of breast cancers

Comparison of the transcriptional signatures of the molecular breast cancer subtypes with gene expression profiles of cell subsets isolated from mammary tissue suggests that different molecular subtypes of breast cancer may arise from cells at various stages of differentiation along the mammary epithelial hierarchy. In other words, molecular subtypes of breast tumors may be reflective of arrest at different stages of epithelial cell development (Figure 1.7). In this scheme, the claudin-low group includes the most primitive tumors that are the most similar to the mammary stem cell, and the next step on the pathway is the luminal progenitor, that seems to be the cell of origin of basal-like molecular subtype of breast cancers. *BRCA1* mutation is linked to this luminal progenitor/basal-like phenotype, and thus, somehow, loss of BRCA1 function may block further differentiation and keep a cell in this step of development. The next step in development is the luminal progenitor/basal-like cell that seems to be the cell of origin of the HER-2–enriched molecular subtype consisting basically in a loss of basal characteristics and a gain of luminal characteristics. The most differentiated groups are the luminal A and luminal B tumors (Prat and Perou, 2009; Perou, 2011). This scheme is somewhat complicated by the possibility that tumor cells may drift from their original configuration, losing specific markers of the cell of origin and taking on others (epithelial plasticity). A prime example of this is EMT, in which transformed epithelial cells switch to a mesenchymal phenotype whose expression profile resembles that of stem cells (Bertos *et al*., 2011).

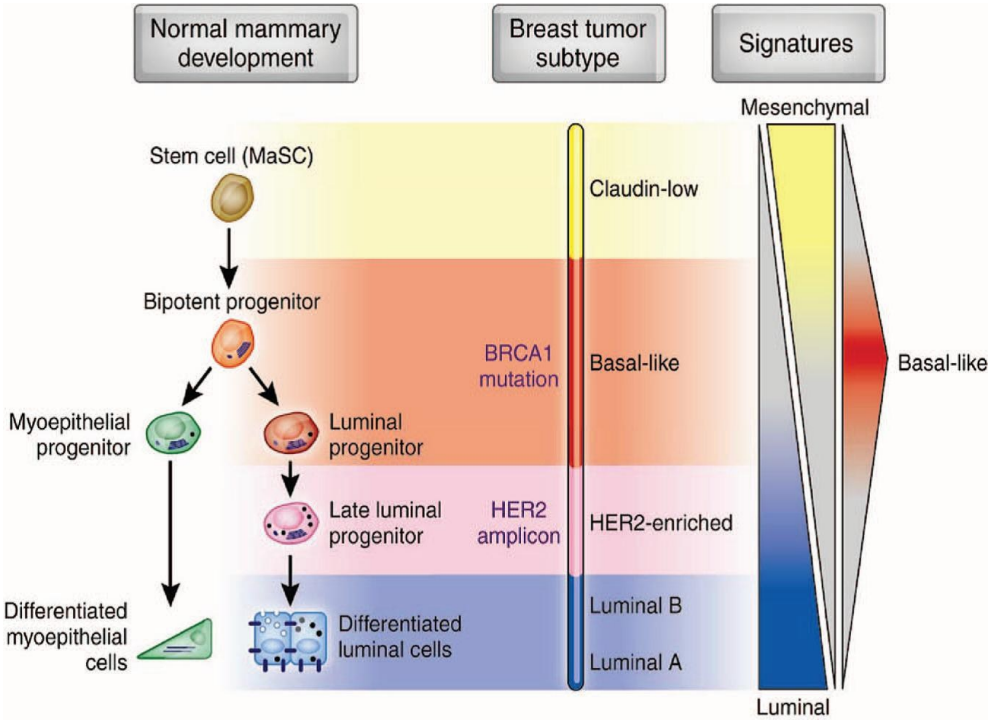


Figure 1.7. Functional classification of breast cancers (Perou, 2011).

1.4.3 Luminal progenitor as cell of origin of BRCA1-mutated breast cancers

Initial pathological and histological studies of *BRCA1* mutant tumors suggested that these cancers may arise from the transformation of a cell type restricted to the basal layer of the breast epithelium (Liu *et al.,* 2008). Recently, a number of publications have argued for the luminal progenitors as the target cells of transformation for *BRCA1* mutated breast cancers (Lim *et al* 2009; Prat *et al*., 2010; Molyneux *et al*., 2010; Proia *et al*., 2011). One explication of the discrepancies between these studies could be that BRCA1 controls differentiation of stem cells into its progeny at multiple different steps and, depending on the degree of loss of *BRCA1* expression, the block could occur at different levels: complete loss of *BRCA1* could result in a block at the stem/early progenitor stage, while partial loss may allow the cells to continue to differentiate up to the luminal progenitors (Ginestier *et al*., 2009; Buckley and Mullan, 2012; Figure 1.8b,c). However, none of the studies to date have proven conclusively that the cell of origin of BRCA1 mutated breast cancers is a luminal progenitor cell. For example, Lim et al. showed that breast tissue from *BRCA1*-mutation carriers had an increased number of luminal progenitors. However, observation of an increased or abnormal pool of cells does not mean that these cells are the initial target for transformation. In fact, it is important to emphasize the distinction between the cell of origin and the cancer stem cell. The cells of origin refer to the original cells that were transformed and therefore represent the cancer-initiating cells, while the cancer stem cells may be the result of their transformation/dedifferentiation ensuring tumor maintenance and propagation. Thus, expanded populations of aberrant progenitor cells observed in many recent studies, could represent either the cells of origin or cancer stem cells. Identification of the cells of origin and the cancer stem cells of BRCA1 related cancers will help in establishing diagnosis and in developing novel therapies in the future.

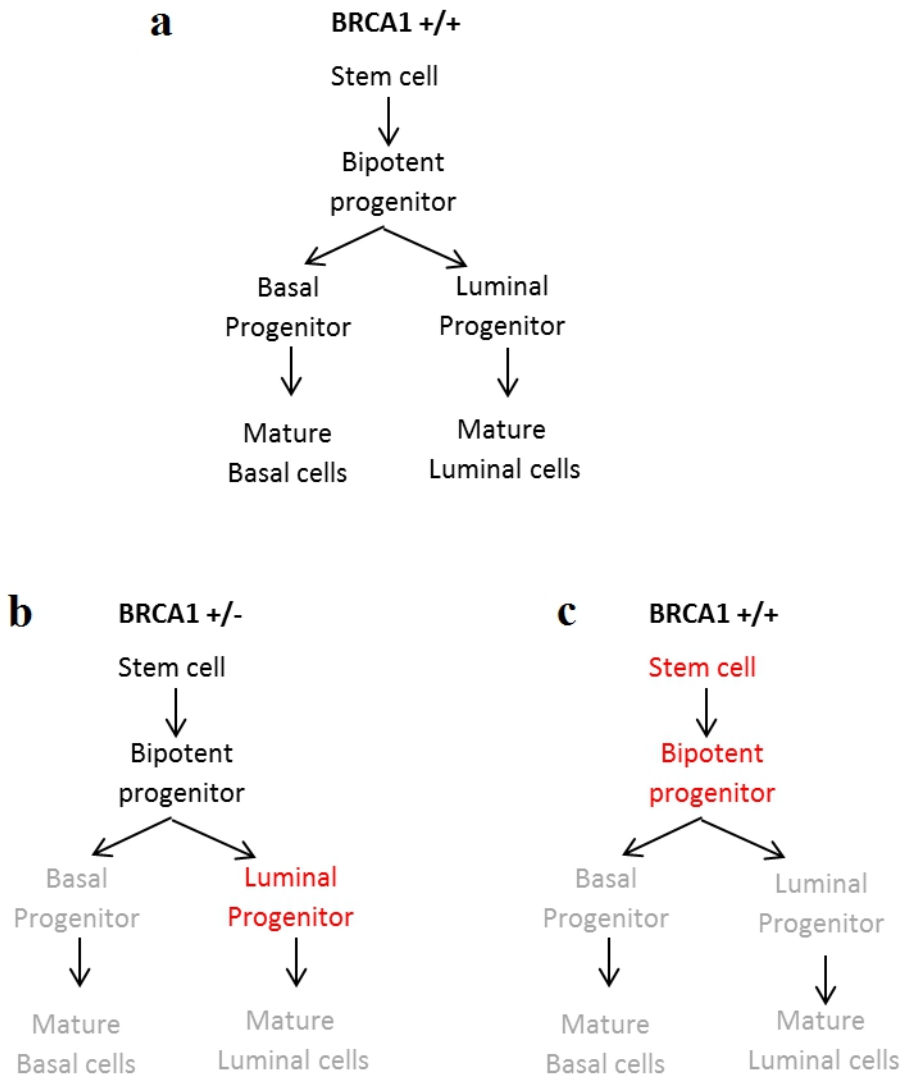


Figure 1.8. Schematic representation of potential outcomes of loss of BRCA1 function on differentiation. (a) Normal mammary gland differentiation. **(b)** Model proposed by Lim *et al*. where heterozygous BRCA1 inactivation results in an increase in luminal progenitors with a block in differentiation down other lineages. **(c)** Model proposed by Liu *et al*. where complete loss of BRCA1 function results in a block in differentiation from the bipotent progenitor and therefore an increase in stem/early progenitor cells (Buckely and Mullan, 2012).

1.5 Resources, tools and models to study breast cancer stem cells

1.5.1 Human breast cancer cell lines

Human breast cell lines have long served as models for a wide array of applications including the study of molecular, cellular, and biochemical mechanisms that regulate breast epithelial biology. Their ability to accurately reflect phenotypes of tumors remains controversial. Several studies have suggested that cell lines exhibit genetic alterations due to adaptation of culture environment, and are poor predictors of *in vivo* sensitivity to drug efficacy (Perou *et al*., 2000; Sorlie *et al*., 2001). However, other studies have indicated that cell lines are appropriate models to study tumor because they mirror many of the biological and genomic properties found within primary human tumors (Bergamaschi *et al*., 2006; Chin *et al*., 2006). Genomic approaches have revealed that, like primary tumors, the gene expression signatures of breast cancer cell lines can distinguish luminal from basal subtypes of breast cancer. Moreover, cell line-derived gene signatures can correctly classify human tumor samples, suggesting that, despite their acquired ability to grow *in vitro* and acquired mutations following adaptation to culture conditions, cell lines continue to share many of the molecular and genetic features of the primary breast cancers from which they were derived (Keller *et al*., 2010). Actually, breast cancer cell lines are widely used as models to study breast cancer stem cells. Recently, based on transcriptional profiling, breast cancer cell lines have been classified in one luminal and two basal-like (A and B) subtypes. Luminal lines resemble luminal A/B tumors, basal-A lines resemble basal-like tumors, and basal-B lines display many stem/progenitor cell characteristics (Kao *et al.,* 2009). I focused my thesis work on basal-A breast cancer cell lines, because they are related to basal-like tumors that are the tumor subtype targeted of my research. Particularly, I hypothesized that the basal-A *BRCA1* mutated breast cancer cell line HCC1937 could be the best candidate for identifying novel cancer stem cell markers. To date, attempts to establish breast cancer cell lines from germ-line mutation

related breast cancers have been mostly unsuccessful. HCC1937 is a *BRCA1* mutated cell line established from a primary non-metastatic IDC originating from a 24-year-old patient with a germ-line mutation. The cell line is homozygous for the *BRCA1* 5382insC mutation, whereas the patient's lymphocyte DNA was heterozygous for the same mutation. HCC1937 also has an acquired mutation of *TP53* with wild-allele loss, and an acquired homozygous deletion of the *PTEN* gene. No significant levels of progesterone or estrogen binding were observed in either the primary tumor or the HCC1937 cultured cells. Only very low levels of *Her-2/neu* were expressed (Tonk *et al*., 1998). Literature data concerning basal-like breast cancers induced by *BRCA1* inactivation led me to hypothesize that these tumors, and likely the corresponding cell lines, might be appropriate to identify CSCs with a phenotype different from the classical $CD44^{+}/CD24^{-/low}$ phenotype. Infact, it has been reported that *BRCA1* basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells (Lim *et al*., 2009; Molyneux *et al*., 2010). Moreover, a recent study reported the existence in the HCC1937 cell line of a stem cell population that was not $CD44^{+}/CD24^{-/low}$ (Hwang-Verslues *et al*., 2009). So, I used the HCC1937 cell line as model to identify novel BCSC markers in basal-like breast tumors.

1.5.2 Mammosphere assay

Cell culture in non-adherent conditions is a widely used tool to identify and isolate breast cancer stem cells. This technique was initially adapted to normal breast obtained from reduction mammoplasty. Human mammary stem and progenitor cells were able to survive in suspension and produce spherical colonies (mammospheres) which were enriched in early progenitor/stem cells and able to differentiate along the three mammary epithelial lineages and to clonally generate complex functional structures in reconstituted 3D culture systems as well as reconstitute human normal mammary gland in mice (Dontu *et al*., 2003). This system

is now widely used to study underlying mechanisms of growth under anchorage-independent conditions, and by extension, to discover pathways implicated in stem/progenitor cells survival. The mammosphere assay was also successfully used to establish long-term cultures enriched in stem/progenitor cells from invasive tumor samples. The mammospheres formed in these conditions were called tumorspheres. They showed an increase in $CD44^{+}CD24^{-//low}$ cells, overexpressed neoangiogenic and cytoprotective factors, expressed the putative stem cell marker OCT4, and displayed high tumorigenic potential in NOD/SCID mice (Ponti *et al*., 2005). Thus, the development of *in vitro* suspension culture systems not only provides an important new tool for the study of mammary cell biology, but also has important implications for understanding key molecular pathways in both normal and neoplastic stem cells. Mammospheres also showed an increase in "side population" (SP) fraction that is linked to a property of stem cells. Particularly, the progenitor/stem cell population have a resistance to environmental toxins, chemotherapeutic and radiation agents, often as a result of multidrug resistance (MDR) via expression of ABC (ATP-Binding Cassette) family of transporter proteins, a property not found in differentiated cells. One of the substrates of transporter proteins is the fluorescent Hoechst 33342 dye. So, stem/progenitor cells actively extruding Hoechst 33342 dye, become negative for this dye, whereas differentiated cells remain positive. The negative, stem-like population is called Side Population (SP).

1.5.3 Expression of surface markers

Another widely used tool for the isolation of cancer stem cells is the expression of surface markers. The pioneering study by Clarke and colleagues used breast cancer xenografts to isolate a population of cells able to initiate tumors in NOD/SCID mice. This population displayed the surface marker phenotype $CD44^+CD24^-$ (Al-Haii *et al*., 2003). Given the small number of tumors analyzed by Clarke's group and the characteristics of these tumors (eight out of nine were pleural effusions), their findings cannot be applied to a larger number of tumors that cover the high diversity of breast cancer subtypes. In a subsequent study, Honeth and colleagues have resolved Al-Hajj's study limitations by using a large collection of well characterized breast cancers in order to explore the prevalence of cells with different CD44/CD24 phenotypes within breast tumor subtypes (Honeth *et al.,* 2008). Only 31% of the tumors analyzed contained $CD44^+CD24^-$ cells, ranging from 1% or less to 100% of the total tumor population, and the $CD44^+CD24^-$ phenotype was more common in basal-like tumors. The findings of Honeth and colleagues suggest that $CD44^+CD24^-$ is the phenotype associated with tumor initiating cells only in certain breast cancers. Furthermore, Hwang-Verslues and colleagues found the existence of a cancer stem cell population that was not $CD44^+CD24^-$ in the cell line HCC1937 (Hwang-Verslues *et al*. , 2009). According to this findings, Meyer and colleagues showed that both $CD44^+CD24^-$ and $CD44^+CD24^+$ cell populations sorted from ER negative breast tumors are tumorigenic in murine xenograft models, indicating that xenograft initiating cells are not always restricted to the $CD44^+CD24^+$ population (Meyer *et al*., 2010). Moreover, recent studies suggest that $CD44^+CD24^-$ antigenic phenotype is an hallmark of EMT rather than stem cells. Particularly, it has been demonstrated that induction of EMT results in the acquisition of stem properties and is associated with the surface expression of $CD44^+CD24^-$ antigenic profile (Mani *et al*., 2008; Morel *et al*., 2008) These and other studies have highlighted the need for additional markers to further discriminate the BCSCs. In the

present study, I explored as novel candidate BCSC markers a series of selected 28 surface antigens including molecules known to be stem cell and/or cancer stem cell markers and molecules involved in processes such as cell adhesion, migration, apoptosis, metastasis, cell signaling, proliferation and differentiation.

1.5.4 Side Population and ABCG2

One of the analyzed surface markers explored in the present study as novel BCSC markers, is the breast cancer resistance protein 1 (BCRP1). BCRP1 is a member of the ATP-binding cassette (ABC) transporter G family and it is also known as ABCG2 or CD338. It is a 655-amino acid polypeptide and contains six putative transmembrane domains. The high *BCRP1* gene expression leads to increased drug efflux and reduced intracellular drug concentration resulting in decreased cytotoxicity. *ABCG2* has been found overexpressed in several human tumors indicating its possible role in multidrug-resistant phenotype of various cancer cells. ABCG2 protein plays an important role in the stem cell research in fact, normal and cancer stem cells overexpress transmembrane transporters, including ABCG2/BCRP1. This protein excludes the fluorescent Hoechst 33342 dye from the cells and so, it seems to be one of the major mediators of side population. The strongest evidence linking ABCG2 and the side population phenotype comes from the nearly complete loss of the bone marrow in *Abcg*$^{-/-}$ mice (Zhou *et al*., 2002). Other supporting evidence is that side population cells preferentially express *ABCG2* (Summer *et al*., 2003; Shimano *et al*., 2003) and that *ABCG2* expression is detected in known stem/progenitor cells such as hematopoietic (Zhou *et al*., 2001) and neural stem cells (Cai *et al*., 2004). On the other hand, it should be noted that only a fraction of the side population cells expresses ABCG2 and that the two side populations and stem/progenitor cells express additional ABC transporters. Moreover, Patrawala *et al*. showed that, in contrast to the tumorigenic differences between the side population and non–side population cells, the

$ABCG2^+$ and $ABCG2^-$ cancer cells show very similar tumorigenic potentials *in vivo* (Patrawala *et al*., 2005).

The SP technique has been used for many years to isolate both normal and cancer stem cells from different organs and species (Hirschmann-Jax *et al*., 2004; Minn *et al*., 2007; Montanaro *et al*., 2004). SP cells isolated from mammoplasty of healthy patients, were found to express more BCRP1 than non-SP cells. Moreover, a specific BCRP1 inhibitor (Ko143) reduced SP formation, suggesting that BCRP1 confers the SP phenotype in mammary epithelial cells. Interestingly, SP cells did express neither luminal nor myoepithelial markers (MUC1/CD227 and CALLA/CD10) nor estrogen receptor (ER), whereas non-SP cells did (Clayton *et al.*, 2004), suggesting that they are dedifferentiated. Cells isolated extract from healthy mammary gland, cultured in non-adherent serum-free culture conditions, and shown to be enriched in stem/progenitors cells, also display a dedifferentiated phenotype (Dontu *et al*., 2003). The SP fraction from uncultured mammary cells represented ~1% of cells. In contrast, in mammospheres, the SP fraction represented 27% of the cells and could generate bi-lineage colonies when cultured under differentiating conditions, suggesting that the SP fraction contained the bipotent progenitors and was capable of mammosphere formation (Dontu *et al*., 2003).

II. RATIONALE AND AIM OF THE THESIS

Breast cancer is the first human carcinoma for which a putative cancer stem cell subpopulation (BCSCs, breast cancer stem cells) has been isolated with the $CD44^+/CD24^-$/low phenotype (Al-Hajj et al., 2003). Given the small number of tumors analyzed by Clarke and colleagues and the characteristics of these tumors (eight out of nine were pleural effusions), their findings cannot be enlarged to the tumors that cover the high diversity of breast cancer subtypes. In a subsequent study, Honeth and colleagues have resolved Clarke's study limitations and demonstrated that $CD44^+CD24^-$ is the phenotype associated with tumor initiating cells only in certain breast cancers (Honeth *et al*., 2008). According to this findings, Meyer and colleagues showed that both $CD44^+CD24^-$ and $CD44^+CD24^+$ cell populations sorted from ER (estrogen receptor) negative breast tumors are tumorigenic in murine xenograft models, indicating that xenograft initiating cells are not always restricted to the $CD44^+CD24^+$ population (Meyer *et al*., 2010).

Some questions arise from these studies: Is the $CD44^+CD24^-$ phenotype associated with tumor initiating cells only in certain breast cancers? Are cancers that do not contain cells with this phenotype driven by a different cancer stem cell? If this is the case, do these cancer stem cells not bearing the $CD44^+CD24^-$ phenotype have a different origin? At this point, many studies are still debating on the relevance of CD44 and CD24 in identifying BCSCs. Thus, there is the need for additional markers to further discriminate the BCSCs.

The central aim of the present thesis is to identify novel markers for isolation and characterization of breast cancer stem cells. To achieve this aim, my work was conducted to reach three objectives:

1) Identification of putative BCSC markers other than CD44 and CD24 in breast cancer cell lines representative of different molecular breast cancer subtypes.

To reach this objective, I conducted two different tasks: (1a) First, I characterized different breast cancer cell lines by flow cytometry, which allowed me evidencing novel putative markers; then, (1b) I isolated from the breast cancer cell lines the putative BCSC subpopulations expressing the markers identified in task 1a).

2) Validation of stem properties and tumorigenicity of the newly identified putative BCSC subpopulations.

(2a) I first conducted *in vitro* experiments to explore stem properties of isolated cell subpopulations; then, (2b) I conducted *in vivo* experiments in NOD/SCID and *nude* mice to explore the tumorigenicity of the putative BCSCs isolated.

3) Standardization of flow-cytomertic methods to study primary cultures from human breast cancer tissues.

Tasks #1 and #2 were accomplished in the first part of the present thesis, and they allowed identifying and isolating a subpopulation of cells with stem cell-like properties in a breast cancer cell line. Task #3 was accomplished in the second part of the present thesis, and it will subsequently be used to isolate the putative BCSC population identified in tasks #1 and #2 from human breast cancer tissues.

While accomplishing task #3, we have also determined whether cell cultures established from human breast cancer tissues maintain the cellular heterogeneity of primary tissues and may therefore be used for *in vitro* modeling of breast cancer subtypes. To this aim, we used a much larger vocabulary of surface markers than those used in previous studies to characterize primary cultures by flow cytometry and we analyzed luminal cytokeratins CK18, CK19, and myoepithelial/basal CK5, SMA (alpha-smooth muscle actin) and vimentin expression by western blot.

III. PART 1:

Identification of a novel BCSC phenotype in the basal-A BRCA1 mutated breast cancer cell line HCC1937

In the first part of the thesis task #1 and task #2 of the central aim have been accomplished.

1) Identification of putative BCSC markers other than CD44 and CD24 on breast cancer cell lines representative of different molecular breast cancer subtypes.

To reach this objective, first, different breast cancer cell lines have been characterized by flow cytometry, allowing to evidence new putative markers in breast cancer cell line HCC1937. Then, the putative BCSC subpopulations expressing the markers of interest have been isolated by cell sorting.

2) Validation of stem and tumorigenic properties of the newly identified putative BCSC subpopulations by *in vitro* and *in vivo* assays.

This work allowed to identify a novel BCSC phenotype, $CD44^{+}CD24^{+}CD338^{bright}$, in the basal-A *BRCA1* mutated breast cancer cell line HCC1937.

3.1 RESULTS

3.1.1 Surface expression profile of breast cancer cell lines.

3.1.1.1 Experimental strategy: cell lines and surface markers studied

In order to identify novel BCSC markers, I started with a wide flow cytometry characterization of several cell lines representative of the different molecular subtypes of breast cancer. I focused my interest on the luminal MCF7, the basal-B Hs578T, and two basal-A cell lines, BT-20 and HCC1937. Table 1 shows the main characteristics of the chosen cell lines. I evaluated the expression of thirty surface antigens that are involved in processes such as cell adhesion, migration, apoptosis, metastasis, cell signalling, proliferation and differentiation, and molecules known to be stem cell and/or cancer stem cell markers in other tumors (Table 2). Figure 3.1 shows the result of surface markers analysis in HCC1937 cell line. I performed flow cytometry (FCM) analysis using a four-color flow FCM panel. In fact, I stained cells with monoclonal antibodies against the two classical breast cancer stem cell markers, CD44 and CD24, combined with pairs of antibodies against molecules explored as potential novel BCSC markers (Table 5, Materials and Methods section). Hence, I could analyze and compare the expression of 28 surface markers in the $CD44^{+}/CD24^{-/low}$ and $CD44^{+}/CD24^{+}$ cell subpopulations of each cell line.

Table 1. Breast cancer cell lines characterized by flow cytometry.

Cell Line	Source	Disease	Receptors	Oncogenes	Cell Line subtype	Tumor subtype
BT-20	P.Br	IDC	TPN		Basal-A	Basal-like
HCC1937	P.Br	DC	TPN	BRCA1	Basal-A	Basal-like
Hs 578 T	P.Br	IDC	TPN		Basal-B	?
MCF-7	PE	IDC	ER+ PR+ HER2-		Luminal	Luminal

DC = Ductal Carcinoma; IDC = Invasive Ductal Carcinoma; ER = Estrogen Receptor; PR = Progesterone Receptor; HER2 = Human Epidermal growth factor Receptor 2; TPN = Triple Negative; P.Br = Primary Breast; PE = Pleural Effusion.

Table 2. Molecular identity and function of the antigens analyzed in breast cancer cell lines.

CD	Molecule	Function	References
CD9	P24	Cell adhesion and migration	Nishida 2009
CD10	CALLA	Antigen overexpressed on many tumors	Iwaya 2002
CD24	HSA	Adhesion and metastatic tumors	Al-Hajj 2003
CD26	DPPIV	Exopeptidase, tissue restructuring	Pro 2004
CD29	β1 integrin	Adhesion to matrix proteins	Charafe-Jauffret 2009
CD44	H-CAM	Cell polarity, suppression of apoptosis, metastasis	Al-Hajj 2003
CD47	Rh-associated protein	Cell activation, apoptosis, cell spreading	Manna 2004
CD49b	α2 integrin	Cell adhesion to collagen and laminin	Langsenlehner 2006
CD49f	α6 integrin	Cell adhesion, migration, cell surface signaling	Cariati 2008
CD54	ICAM	Cell adhesion, immune reactions	Rosette 2005
CD55	DAF	Protection against complement	Ikeda 2008
CD59	MIRL	Protection from complement-mediated lysis	Babiker 2005
CD61	β3 integrin	Cell adhesion, cell signaling	Charafe-Jauffret 2009
CD66b	CEACAM8	Cell adhesion, cell signalling	Lasa 2008
CD66c	CEACAM6	Cell adhesion, cell signalling	Lasa 2008
CD81	TAPA-1	Response to antigens	Yáñez-Mó 2009
CD90	Thy-1	Cell adhesio and differentiation	Yamazaki 2009
CD105	ENG (Endoglin)	Angiogenesis, vessel wall integrity	Henry 2011
CD133	Prominin 1	Unknown	Wright 2008
CD151	PETA-3	Cell adhesion	Sadej 2009
CD164	MGC-24	Adhesion and homing	Havens 2006
CD165	AD2	Unknown	Seon 1984
CD166	ALCAM	Adhesion, organ development	Kulasingam 2009
CD184	CXCR4	Increased expression in mammospheres	Krohn 2009
CD200	OX2	Immunosuppression	Kawasaki 2007
CD227	MUC1	Response to hormones and cytokines	Stingl 2009
CD324	E-Cadherin	Cell adhesion, tumor suppression	Prasad 2009
CD326	EpCAM	Cell adhesion	Stingl 2009
CD338	ABCG2	Pumping cytotoxic drugs out of cells	Zhou 2001
CD340	Her2/neu	Cell growth and differentiation	Paik 2008

3.1.1.2 Surface expression profile of HCC1937 basal-A cell line.

To evaluate the differences of surface markers expression between $CD44^+/CD24^{-/low}$ and $CD44^+/CD24^+$ subpopulations of each cell line, I determined the ratio between the mean fluorescence intensity (MFI, see section 6.2 of the Flow Cytometry Appendix) of each CD in $CD44^+/CD24^+$ and $CD44^+/CD24^{-/low}$ cells. This analysis didn't show any significant differences of surface markers expression between $CD44^+/CD24^{-/low}$ and $CD44^+/CD24^+$ subpopulations of MCF7, Hs578T and BT-20 cell lines. Contrarily, there were some interesting differences in the HCC1937 basal-A cell line. Table 3 shows the ratio of MFI of each antigen evaluated in the two cell subpopulations, $CD24^+$ and $CD24^-$, present in the HCC1937 cell line. Most CDs displayed a ratio near to 1 which indicated an equal or very similar expression in the two cell subpopulations, whereas CD338/ABCG2 and CD10/CALLA displayed a ratio greater than two (2.4) which indicated that they are expressed at a higher level in the $CD24^+$ than in $CD24^-$ cell subpopulation. This observation suggested that, in the HCC1937 basal-like breast cancer cell line, some stem cell markers are highly expressed in the $CD44^+/CD24^+$ subpopulation. Particularly, CD338, that has been reported to be highly expressed in stem cells (Zhou *et al*., 2001; Cai *et al*., 2004), displays a complex expression in the HCC1937 cell line (Figure 3.2). Indeed, some HCC1937 cells expressed CD338 at low level, whereas others expressed it at high level. Strikingly, the MFI of CD24 in cells overexpressing CD338 was 4.7 fold higher than in cells expressing CD338 at low level (5,200 and 1,100 respectively; $p<0.05$). This result indicates that CD338 overexpressing cells also do overexpress CD24. I thus reasoned that HCC1937 cell line could contain a putative $CD24^+CD338^+$ CSC subpopulation.

Table 3. Analysis of surface antigens expression in $CD24^{hi}$ and $CD24^{-/low}$ cell subpopulations of HCC1937 cell line.

Antigen (CD)	MFI in $CD24^{hi}$ cells	MFI in $CD24^{hi}$ cells	R_{MFI}	p-value
49f	35,075.33	26,455.33	1.3	0.1065
338	3,053	1,291.75	2.4	0.0009
90	5,369.33	3,379.33	1.6	0.0609
133	1,654	1,358.67	1.2	0.4104
47	10,610	8,321.33	1.3	0.3071
200	1,541.67	1,121.33	1.4	0.1801
227	7,171.33	4,982.33	1.4	0.1003
10	1,877	764,5	2.4	0.0129
324	7,075.33	4,192.33	1.7	0.1451
29	78,047,33	63,620	1.2	0.0522
66c	13,176	8,175.33	1.6	0.1536
61	1,266.67	1,195.33	1.0	0.4328
164	5,068.33	3,419	1.5	0.0323
340	4,105	2,653.67	1.5	0.1489
49b	98,598.33	70,022.33	1.4	0.0505
184	2,042	1,100	1.8	0.2849
9	34,267	20,817.33	1.6	0.1232
54	35,329	22,530.4	1.6	0.0970
26	3,305.67	2,486	1.3	0.0134
55	57,866.33	54,469	1.1	0.7307
81	2,4084.2	22,350.2	1.1	0.0084
151	5,578	3,677	1.5	0.0117
66b	2,012	1,566.33	1.3	0.0585
59	5,868	3,682.33	1.6	0.0246
165	2,406	1,686	1.4	0.0339
166	4,753.5	3,255	1.5	0.3157
326	40,513.67	33,192	1.2	0.0640
105	2,859	2,228.33	1.3	0.0630

MFI = Mean Fluorescence Intensity. R_{MFI} = MFI in $CD24^{hi}$/MFI in $CD24^{-/low}$. Each measure was carried out 3-5 times. Most CDs display a R_{MFI} near to 1 indicating an equal or very similar expression in the two cell subpopulations, whereas CD338 and CD10 display a ratio greater than two indicating that they are higher expressed in the $CD24^{+}$ than $CD24^{-/low}$ cell subpopulation ($p = 0.0009$ and 0.0129, respectively).

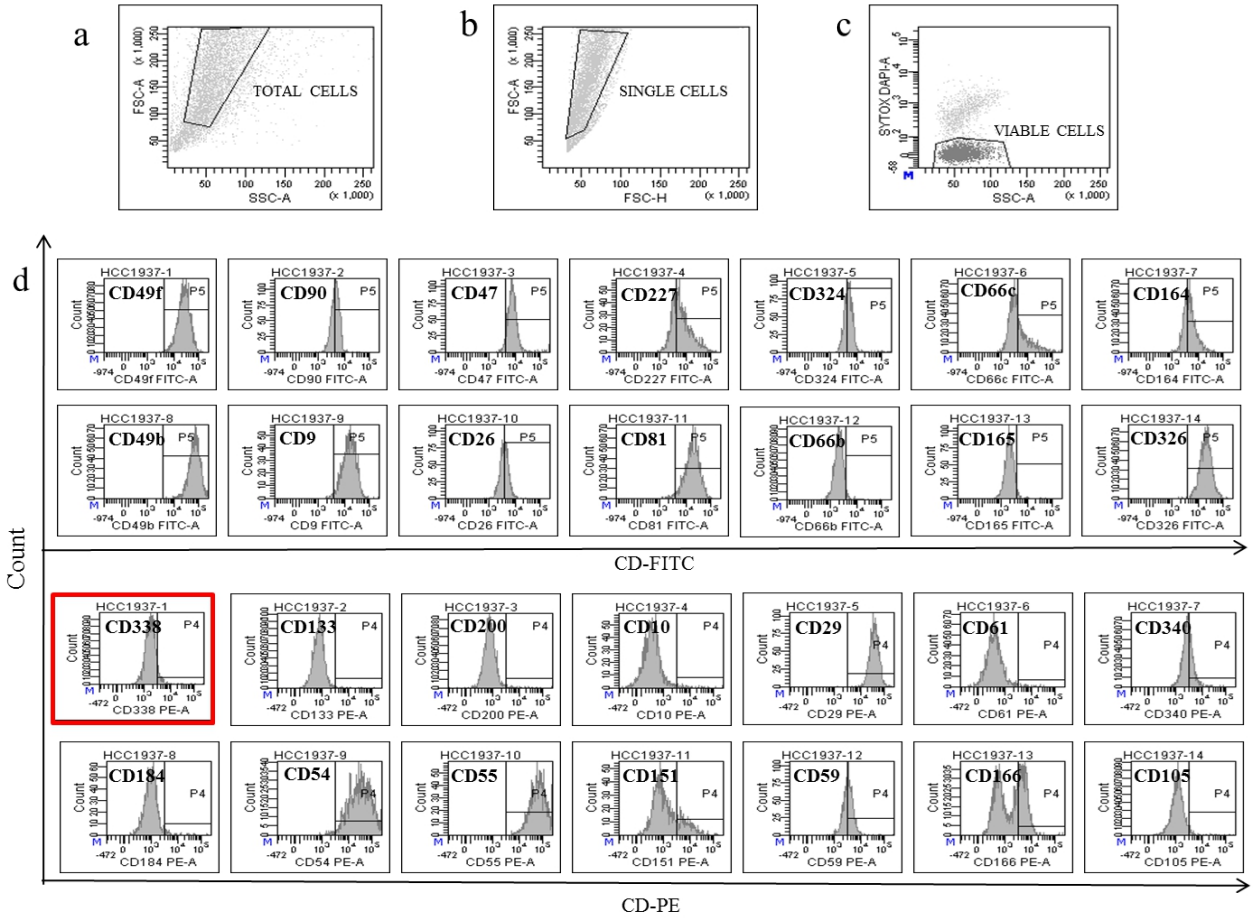


Figure 3.1. Flow cytometry analysis of surface markers expression in the HCC1937 cell line. HCC1937 cell line was prepared for flow cytometry analysis as described in Materials and Methods. For each tube 10,000 events were recorded and analyzed. **(a,b,c)**: gating strategy used to remove debris, cell clusters and dead cells. **(a)** In order to exclude dead cells and debris, cells were gated on a two physical parameters dot plot measuring forward scatter (FSC) *vs* side scatter (SSC) (see section 6.2 of Flow Cytometry Appendix); **(b)** then doublets were excluded by gating cells on FSC-Height *vs* FSC-Area dot plots (see section 6.2 of Flow Cytometry Appendix); **(c)** finally, in order to further exclude dead cells, I gated Sytox Blue negative cells. This gating strategy was used to define the cell population subsequently analyzed for the expression of surface markers. **(d)** Surface markers expression on cells gated as described. The expression of each antigen is represented on a frequency distribution histogram (count *vs* FITC or PE signal, see section 6.2 of Flow Cytometry Appendix). The vertical marker on each histogram used to detect the antibody-positive cells was established using the appropriate negative controls.

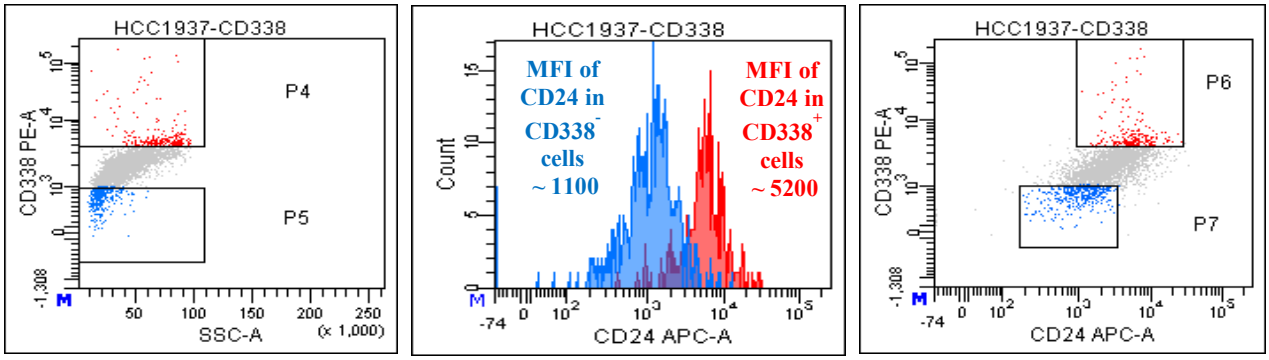


Figure 3.2. Expression of CD338/ABCG2 in the HCC1937 cell line. $CD338^{+}$ cells (panel a, red events) as compared to $CD338^{-}$ cells (panel a, blue events) displayed a mean fluorescence intensity (MFI) of CD24 4.7 fold higher ($p<0.05$), as shown by monoparametric histogram of CD24 expression (panel b) and by CD338 *vs* CD24 dot plot (panel c).

3.1.2. Cell sorting of different cell subpopulations from HCC1937 basal-A cell line

Flow cytometry analysis led me to hypothesize that HCC1937 cells include a subpopulation of stem/progenitor cells overexpressing CD338 and CD24. To validate this hypothesis, I isolated with the cell sorting technique (see section 6.3 of Flow Cytometry Appendix) subpopulations of the HCC1937 cell line expressing or not the markers of interest, CD24 and/or CD338, and compared their stem cell-like and tumorigenic properties through various assays.

3.1.2.1 Isolation of CD338 positive and negative cells and improvement of cell sorting purity

The development of a monoclonal antibody, named 5D3, specifically reacting with the human ABCG2 protein on the cell surface, has previously been reported to successfully sort out viable ABCG2-expressing cells (Zhou *et al*., 2001). 5D3 was thus used as a tool to sort CD388 positive and negative populations out of HCC1937 cells. 5D3 binding to its extracellular epitope strongly depends on the conformation of ABCG2 (Ozvegy-Laczka *et al*., 2005), making the binding particularly unstable. For this reason, in the first part of the study the purity of sorted subsets was not as good as expected (Figure 3.4).

Due to this technical limit, I tried to stabilize the binding of anti-CD338 antibody in order to obtain an improvement of cell sorting purity. To this aim I used, during the cell sorting procedure, a cross-linker protein (PMPI, p-Maleimidophenyl isocyanate) known to be able to increase the 5D3 binding to ABCG2. The PMPI cross-linker protein is able to fix the epitope on a monomeric ABCG2 protein, without disrupting its activity. Figure 3.3 shows the structure of ABCG2, the external epitope recognized by 5D3 and the stabilizing activity of cross-linker proteins. The use of PMPI led to an improvement of sorting purity which increased from 50% to 90%.

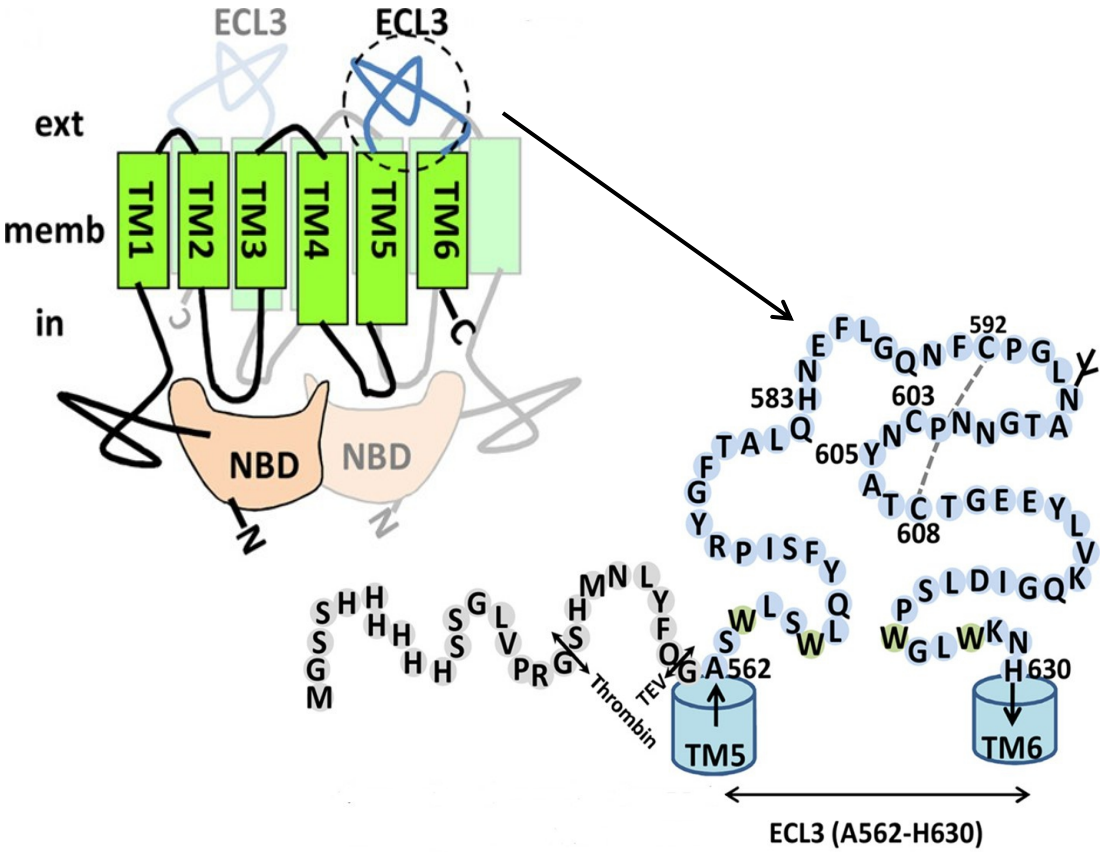


Figure 3.3. ABCG2 structure and cross-linker protein activity. Because ABCG2 is a half-transporter, bearing only one of each of the characteristic ABC family domains (the ATP-binding domain and transmembrane domain), it has to form a homodimer to become functionally active. The ABCG2 homodimer is covalently linked via a disulfide bond formed by cysteines at position 603, localized in the large 55-amino acid-long third extracellular loop (ECL3) of the protein. In ECL3, ABCG2 has two other cysteines at positions 592 and 608. These two residues are indicated as forming an intramolecular disulfide bridge that influences plasma membrane targeting and substrate specificity of the transporter. It seems that activity of cross-linker proteins at the level of cysteines 592 and 608, is responsible for 5D3 increasing binding (Desuzinges-Mandon *et al*., 2010).

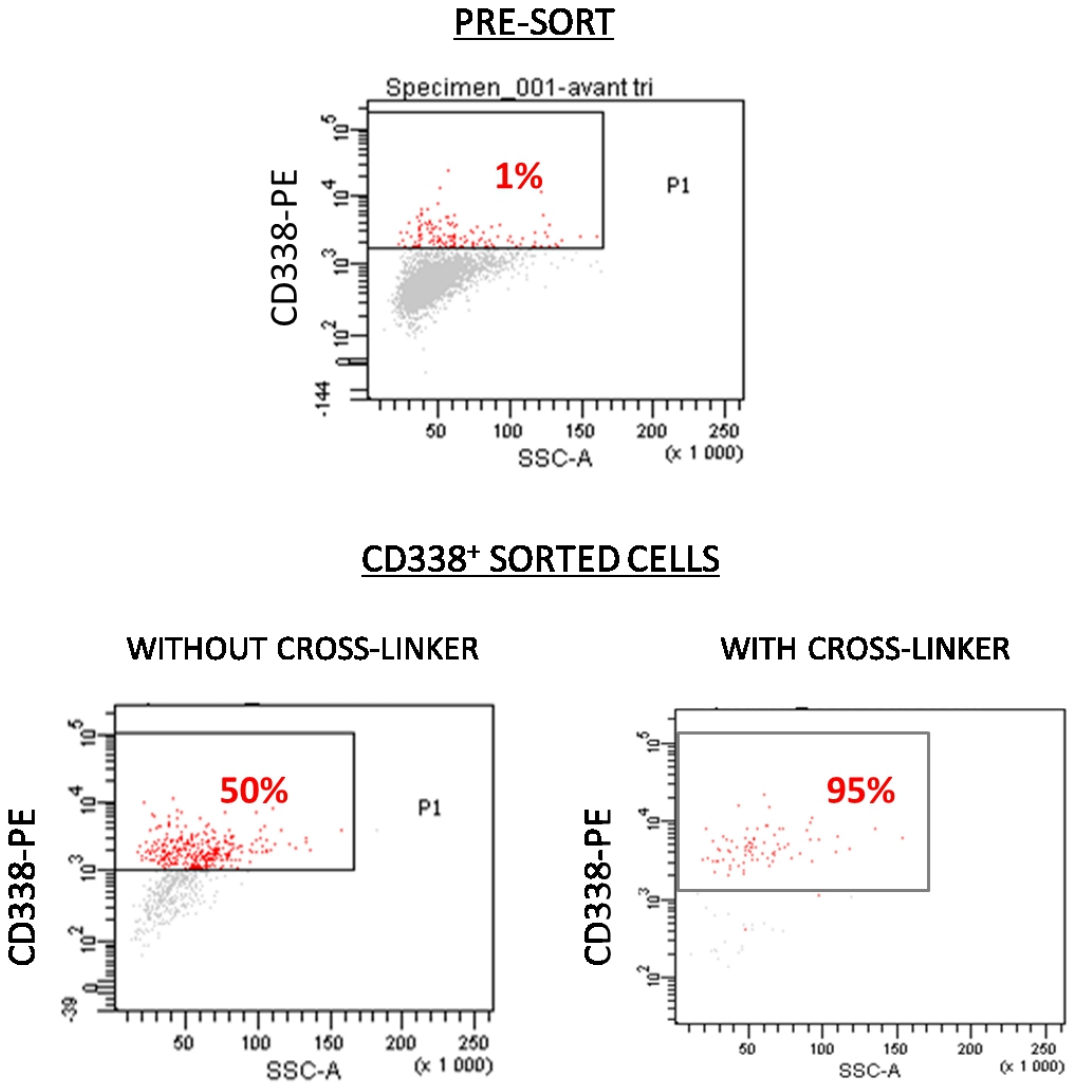


Figure 3.4. Purity of CD338 positive and negative sorted subsets. 5D3 antibody binding to its epitope strongly depends by ABCG2 conformation, so it is not stable. For this reason, in the first part of the study the purity of CD338 positive sorted subsets was low (50%, lower left panel), with a contamination by negative cells in CD338 positive sorted subset. The use of the PMPI cross-linker protein during the cell sorting procedure, increased 5D3 antibody binding leading to an improvement of cell sorting purity (95%, lower right panel).

3.1.2.2 Highlight on CD338 expression in HCC1937 cell line and cell sorting of three different CD338 cell subpopulations

In the initial phase of the study, expression analysis and cell sorting experiments were performed using a BD FACSAria I Cytometer (see Materials & Methods and Flow Cytometry Appendix). This cytometer use the blue laser to discriminate the different fluorescence intensities of PE (Phycoerythrin) that is the fluorochrome conjugated to the 5D3 antibody used in the present study. This laser is able to discriminate just CD338 positive and CD338 negative subsets (+ and – respectively in Figure 3.5, upper left panel). These two subsets were sorted out from HCC1937 cell line in the initial phases of the study. I next had the opportunity to analyze the expression of CD338 with a BD LSR II four-laser flow cytometer. This cytometer is equipped also with another laser, the yellow-green laser, that discriminates the different PE fluorescence intensities more accurately than the blue laser (see section 6.4 of Flow Cytometry Appendix). Taking benefit of this cytometer, a third cell subpopulation expressing an intermediate level of CD338 could be highlighted. HCC1937 cells are thus constituted of three populations 1) $CD338^{+/bright}$ expressing CD338 at high level and consisting just in 1% of the total cell line; 2) $CD338^{neg}$ not expressing CD338 and representing about 20% of the total cell line; 3) $CD338^{+/low}$ expressing CD338 at an intermediate level and consisting in about 79% of the total cell line (red, yellow and blue events, respectively in Figure 3.5, upper right panel). As the LSR II Cytometer is unfortunately an analyzer, but not a cell sorter, cell sorting had to be performed with the FACSAria cell sorter. Nonetheless, I attempted to sort the three populations on the base of the comparison between the two images given by blue and yellow-green lasers of LSR II. As showed in figure 3.5, there was a strong overlap between CD338 negative and low cells analyzed with the blue laser (upper left panel). Consequently, the purity of these two sorted subpopulations was not fully reliable. Figure 3.5 shows an example of cell sorting of bright,

low and negative cell subsets performed with the FACSAria I cell sorter. To better define the level of contamination between the three different populations after cell sorting, I looked at *ABCG2* expression performing a RT-qPCR with sorted cells, using as control the unsorted total cell line. As shown in figure 3.6, $CD338^{bright}$ sorted cells displayed the highest level of ABCG2 mRNA (2.09), whereas CD338 negative sorted cells displayed the lowest one (0.33). $CD338^{low}$ sorted cells displayed an expression level that was intermediate between CD338 negative sorted cells and the unsorted total cell line (0.58). These results indicate that CD338 cell sorting, based on the different ABCG2 surface protein expression, led to the separation of the three subpopulations with different *ABCG2* gene expression levels, i.e. three different cell types. However, the mRNA level difference between $CD338^{low}$ and $CD338^{neg}$ sorted subpopulations is very low, thus confirming a contamination by $CD338^{low}$ cells in $CD338^{neg}$ sorted subpopulation.

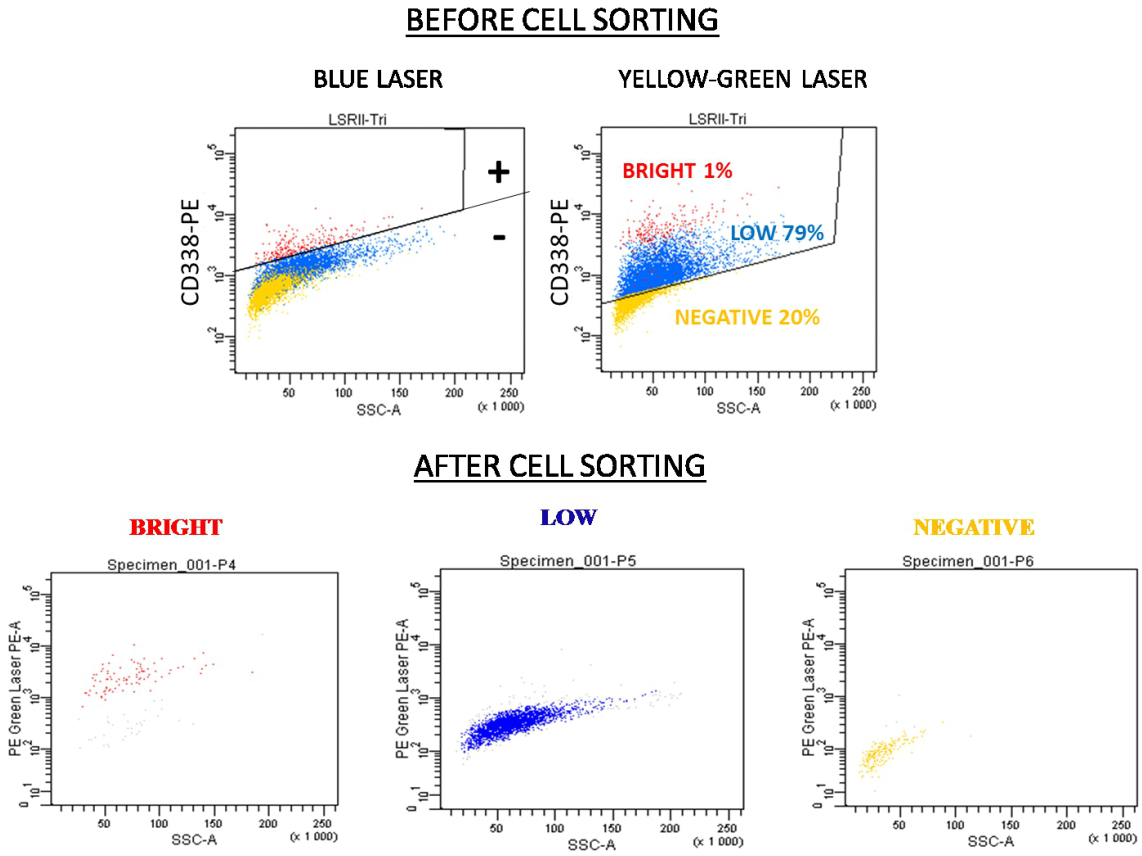


Figure 3.5. Expression of CD338 in HCC1937 cell line and cell sorting of three cell subsets. Upper panels: FACSAria I Cytometer is equipped with the blue laser that is able to discriminate just CD338 positive (upper left panel, red events) and CD338 negative (upper left panel, blue and yellow events). LSR II Cytometer is equipperd also with the yellow-green laser that is able to discriminate three different cell subsets: bright, low and negative cells (upper right panel). Blue events, that seemed to be CD338 negative cells according to the blue laser, actually are CD338 positive (as shown by the yellow laser), although expressing CD338 at lower level than $CD338^{bright}$ cells. Lower panels: example of sorting of the three subsets performed with FACAria I sorter on the base of the images given by the two lasers of LSR II analyzer.

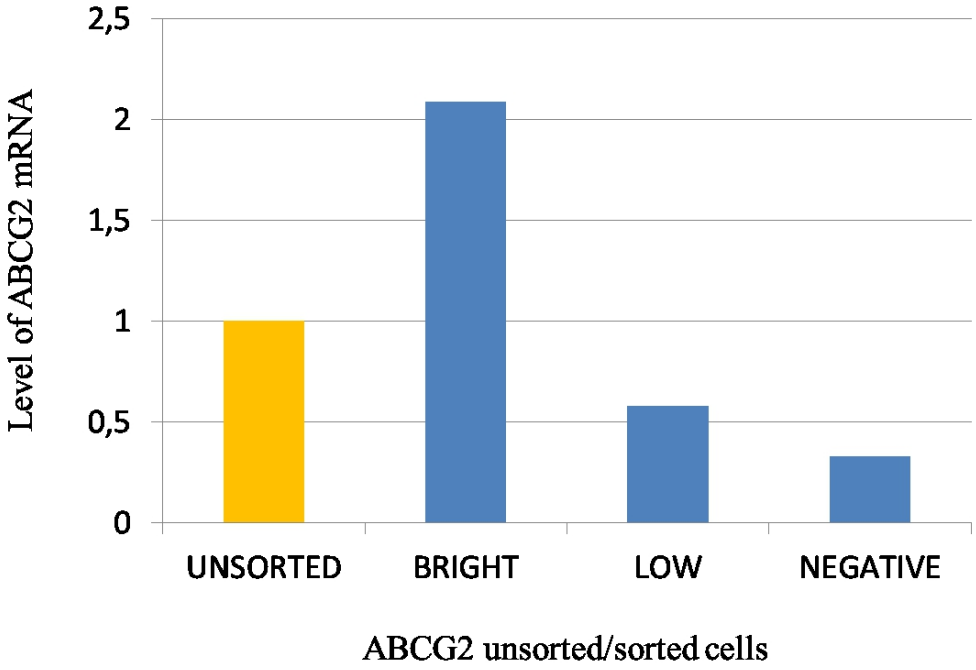


Figure 3.6. *ABCG2* expression analysis in sorted cell subpopulations. Relative mRNA expression levels of *ABCG2* in CD338^{bright}, CD338^{low} and CD338^{neg} sorted populations as estimated by qPCR. The level of mRNA expression in the parental cell line (Unsorted) was set = 1.0 for reference purposes. *HPRT1* was used as a reference gene.

3.1.3 Mammosphere assay with unsorted HCC1937 cell line

Mammosphere forming assay, consists in the evaluation of the ability of cells to grow in non-adherent conditions and to establish structures named as mammospheres, a property considered as a hallmark of mammary stem/progenitor cells. I first performed mammosphere formation assay with unsorted cells and next examined the antigenic expression of cells that constitute the established mammopheres (Figure 3.7). Third-generation mammospheres were disaggregated to a single cell suspension that was subsequently analyzed by flow cytometry. The percentage of $CD338^{bright}$ cells in mammospheres was 4.9 fold higher than that in adherent cell line (mean ± SEM: 8.300% ± 0.5508 and 1.750% ± 0.0866 respectively, $p<0.0001$). Furthermore, the MFI of CD338 in mammosphere-derived cells was 3.8 fold higher than that in adherent cell line. This observation strongly supports the assumption that CD338 positive cells display some stem cell-like properties. It's also interesting to note that all cells from the adherent HCC1937 cell line express CD326/EpCAM and CD49f/α6-integrin (Figure 3.7, panel c), whereas most cells from mammospheres lose the expression of these two markers (Figure 3.7, panel f, gray events). Particularly, in HCC1937-derived mammospheres just $CD338^{bright}$ cells continue to express CD326/EpCAM and CD49f/α6-integrin (Figure 3.7, panel f, red events). CD326 and CD49f delineate distinct subpopulations in human mammary tissue and it has been demonstrated that the double positive subset consists of luminal progenitors (Lim *et al*. 2009). The fact that, in HCC1937-derived mammospheres, just CD338 positive cells continue to express CD326/EpCAM and CD49f/α6-integrin could suggest a progenitor nature for these cells.

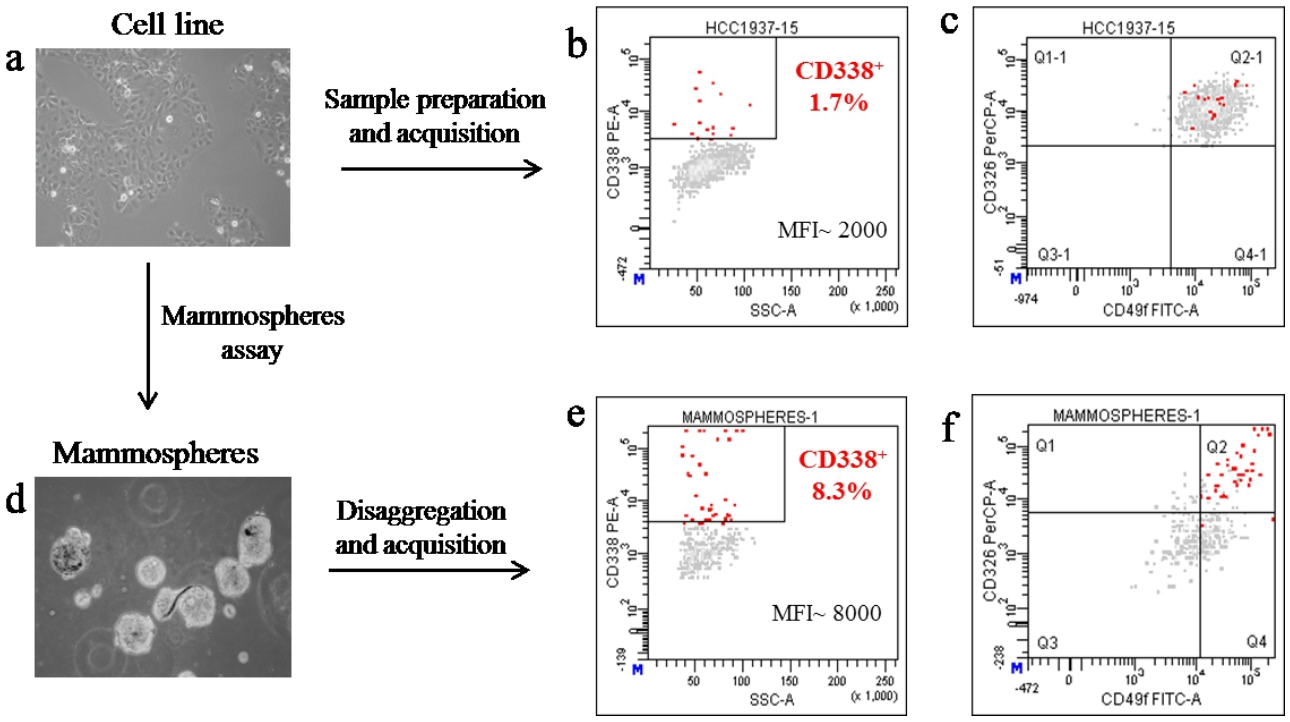


Figure 3.7. Flow cytometry characterization of mammospheres originate from unsorted HCC1937 cell line. (a, b, c): expression of CD338, CD326 and CD49f in cells grown in adherent conditions. **(d, e, f)**: expression of CD338, CD326 and CD49f in cells obtained from the disaggregation of mammospheres. Mammospheres were generated from the HCC1937 cell line as described in Materials & Methods. Third-generation mammospheres were disaggregated to a single cell suspension that was subsequently analyzed by flow cytometry. The percentage of $CD338^+$ cells in mammospheres **(e)** is 4.9 fold higher (mean ± SEM: 8.300% ± 0.5508 and 1.750% ± 0.0866 respectively, $p<0.0001$) than that in adherent cell line **(b)**. The MFI of CD338 in mammospheres-derived cells is 3.8 fold higher than that in adherent cell line. All cells from adherent cell line express CD326 and CD49f **(c)** whereas most cells from mammospheres, except $CD338^+$, lose these two markres **(f)**.

3.1.4 Filiation between the different CD338 sorted cell subpopulations

To evaluate whether CD388^{bright} cells give rise to CD388 low cells, cells were sorted out and their antigenic phenotype was examined after culturing them for several weeks. Particularly, I separately cultured sorted CD338^{bright} and CD338^{low} populations and I analyzed the expression of CD338 after four weeks. While the antigenic phenotype of CD388 low cells remains stable and homogeneous, CD388^{bright} cells gave rise to bright and low cells (Figure 3.8). These results suggested a filiation from CD338^{bright} to CD338^{low} cells. The weak purity of CD388 negatively sorted out cell population avoided us performing similar experiments with this subpopulation. From this result, we can conclude that in the HCC1937 cell line CD338^{low} population could constitute the progeny of the most immature CD338^{bright} population.

.

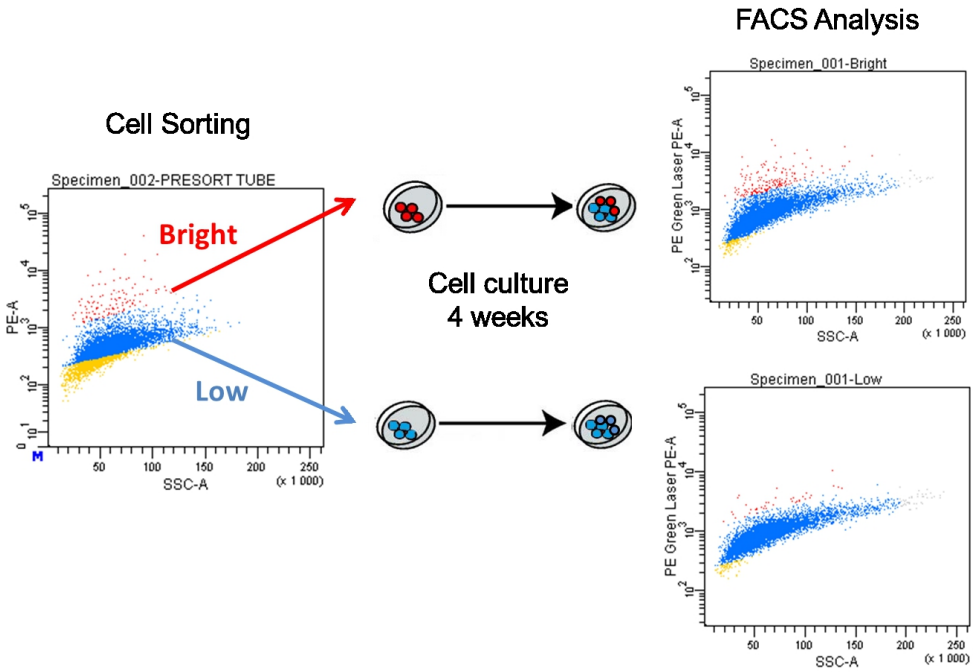


Figure 3.8. Expression of ABCG2 in CD338^{bright} and CD338^{low} sorted cell subpopulations. CD338^{bright} and CD338^{low} cells were sorted out from HCC1937 cell line and plated in adherent conditions. After 4 weeks the two subcultures were detached and the expression of CD338 was analyzed by flow cytpmetry.

3.1.5 Side population in HCC1937 cell line

Progenitor/stem cells are able to actively extrude the fluorescent Hoechst 33342 dye. They are consequently not stained with that dye, while differentiated cells are. The negative-Hoechst 33342 cell population is named Side Population (SP) and ABCG2 seems to be one of the major mediators of side population. To study the side population in the HCC1937 cell line, I used the method described by Goodell *et al*. (see Materials and Methods). Strikingly, HCC1937 cell line includes "side population" representing 1.4% of the total cell line, a percentage very similar to that of $CD338^{bright}$ cells (Figure 3.9, left panel). Furthermore, treatment of HCC1937 with an inhibitor of ABCG2 (Reserpine) reduces the side population from 1.4% to 0.2%, supporting the concept that ABCG2 contributes to the dye exclusion in this cell population. This result indicated that, in the HCC1937 cell line there is an overlap between side population and $CD338^{bright}$ cells, supporting the hypothesis that $CD338^{bright}$ cells are stem/progenitor cells.

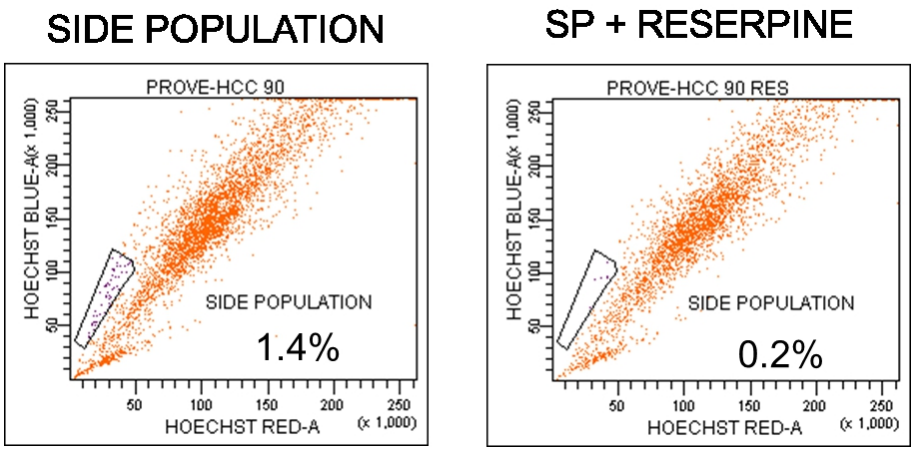


Figure 3.9. Side Population in HCC1937 cell line. The method described by Goodell *et al*., was used to study side population in HCC1937 cell line. In cells treated with the ABCG2 inhibitor Reserpine the percentage of "side population" decreased from 1.4% (left panel) to 0.2% (right panel).

3.1.6 Mammosphere assays with cell subsets sorted out of HCC1937 cell line

To strengthen the hypothesis that stemness properties are specifically allotted to the $CD24^{hi}$ and/or $CD338^{bright}$ cells in HCC1937 cell line, the corresponding populations were sorted out and their mammosphere formation capability was compared to CD24 or CD338 negative couterparts.

3.1.6.1 Mammosphere-forming efficiency of $CD24^{hi}$ and $CD24^{-}$ sorted cells

To compare stem properties of $CD24^{hi}$ and $CD24^{-}$ cells, I performed three independent mammosphere assay experiments from sorted cell subpopulations. In particular, cells from HCC1937 cell line in adherent conditions were stained with anti-CD24 mAb and analyzed by flow cytometry. $CD24^{-}$ and $CD24^{hi}$ cells were separated through cell sorting and then plated in non-adherent conditions at a low density to generate mammospheres (Figure 3.10). The mammosphere formation assay (MFE) of $CD24^{hi}$ cells was higher than that of $CD24^{-}$ cells (mean ± SEM: 5.750 ± 1.031 and 0.500 ± 0.2877 respectively; $p<0.0087$). This result supports the assumption that stem cell-like properties are specifically allotted to the CD24 positive cell subpopulation.

3.1.6.2 Flow cytometry characterization of mammospheres from $CD24^{hi}$ sorted cells

Cells in human mammary tissue can be divided in three distinct subpopulations based on the expression of two surface markers, CD326 (EpCAM) and CD49f (α6-integrin). The $CD49f^{hi}/CD326^{+}$ subset consists of luminal progenitors (Lim *et al*., 2009; Keller *et al*., 2010). Mammospheres from $CD24^{hi}$ sorted cells were disaggregated to single cell suspensions that were then labeled with antibodies against CD338, CD326 and CD49f. The percentage of $CD338^{+}$ cells in mammospheres generated from $CD24^{hi}$ cells is 4.2 fold higher than that in mammospheres generated from $CD24^{-}$ cells (mean ± SEM: 9.667% ± 2.002 and 2.300%

±1.050, respectively; $p<0.05$). Furthermore, the MFI of CD338 in $CD24^{hi}$ mammospheres-derived cells is 2.9 fold higher than that in $CD24^{-}$ mammospheres-derived cells (Figure 3.10). Most cells from $CD24^{hi}$ mammospheres were found to be CD326 and CD49f negative, with the exception of the $CD338^{bright}$ cells that were positive for both surface markers. This phenotype, namely double positivity for CD326 and CD49f, is a marker of "luminal progenitors". I thus concluded that the $CD338^{+}$ cell subpopulation likely displays luminal progenitors phenotype.

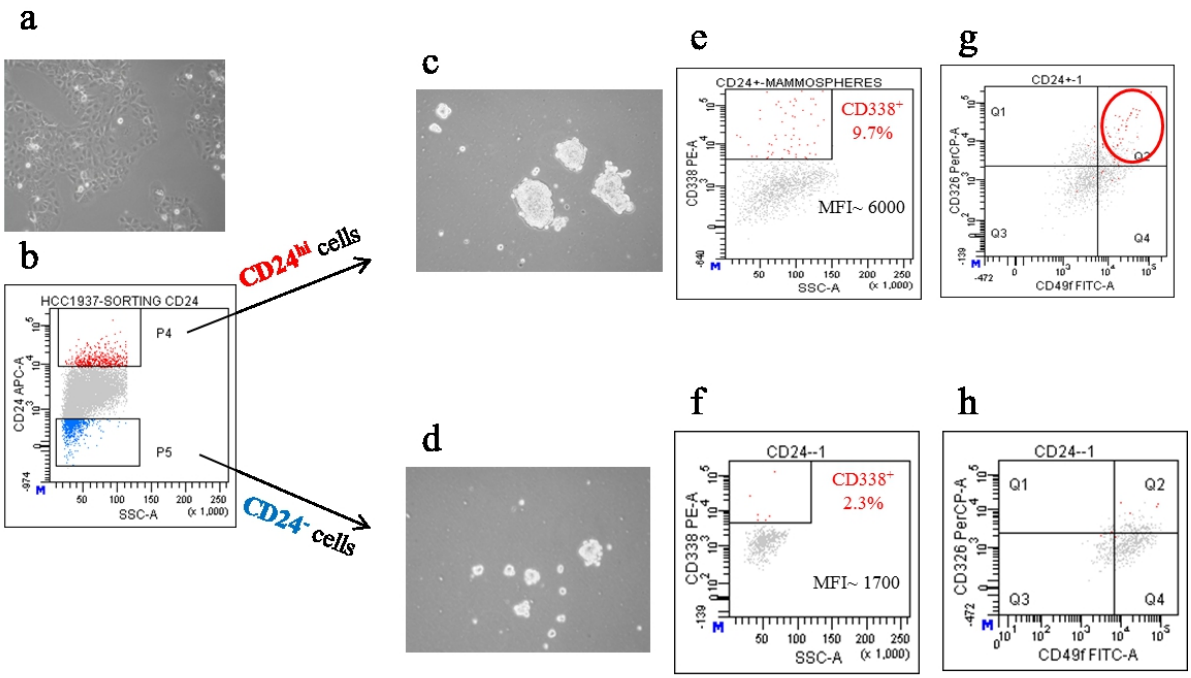


Figure 3.10. Flow cytometry characterization of mammospheres from $CD24^{hi}$ sorted cells. Cells from cell line in adherent conditions **(a)** were stained with anti-CD24 mAb and analyzed by flow cytometry **(b)**. $CD24^{-}$ and $CD24^{hi}$ cells were separeted through cell sorting and plated in non-adherent conditions at a low density to generate mammospheres **(c, d)**. $CD24^{hi}$ cells were able to form mammosphere with a higher efficiency than $CD24^{-}$ cells (mean ± SEM: 5.750 ± 1.031 and 0.500 ± 0.2877 respectively; $p<0.0087$). Mammospheres generated from $CD24^{hi}$ and $CD24^{-}$ cells were disaggregated to single cell suspensions that were then stained with anti-CD338 **(e, f)**, anti-CD326 and anti CD49f **(g, h)** MoAbs. The percentage of $CD338^{+}$ cells in mammospheres generated from $CD24^{hi}$ cells is 4.2 fold higher than that in mammospheres generated from $CD24^{-}$ cells (mean ± SEM: 9.667% ± 2.002 and 2.300% ±1.050, respectively; $p<0.05$) **(e, f)**. The MFI of CD338 in $CD24^{hi}$ mammospheres-derived cells is 2.9 fold higher than that in $CD24^{-}$ mammospheres-derived cells. Most cells from $CD24^{hi}$ mammospheres, but not $CD338^{+}$, lose expression of CD326 and CD49f.

3.1.6.3 Mammosphere-forming efficiency of CD24^{hi}/CD338^{bright} and CD24^{hi}/CD338^{-} sorted cells

Next, I performed a two color cell sorting, that means I stained cells with monoclonal antibodies against both CD24 and CD338 and then separated CD24 positive population in CD338 positive and CD338 negative. Among the CD24 positive cells, those overexpressing the stem cell marker CD338 were able to form mammospheres with higher efficiency than CD338 negative cells (Figure 3.11; mean ± SEM: 13.0 ± 1.080 and 1.5 ± 1.190 respectively; $p<0.05$). This information corroborates the hypothesis that CD24^{+}CD338^{+} cells display stem cell-like properties.

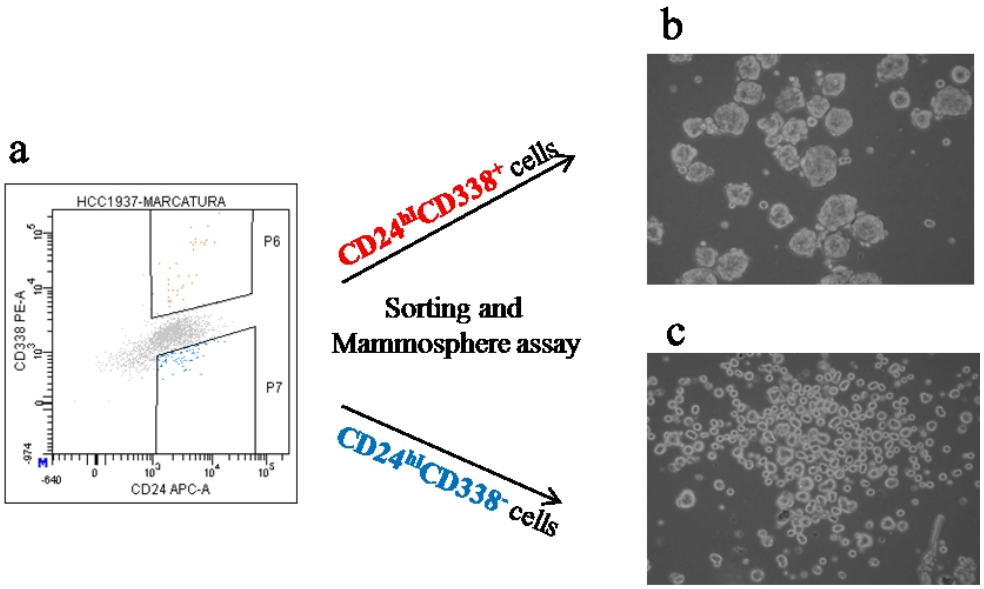


Figure 3.11. Mammosphere-forming efficiency of CD24^{hi}/CD338^{bright} and CD24^{hi}/CD338^{-} sorted cells. Cells from cell line in adherent conditions were stained with anti-CD24 and CD338 MoAbs and analyzed by flow cytometry **(a)**. CD24^{hi}/CD338^{bright} and CD24^{hi}/CD338^{-} cells were separeted through cell sorting and plated in non-adherent conditions at a low density to generate mammospheres **(b, c)**. Among the CD24 positive cells, those overexpressing the stem cell marker CD338 were able to form mammospheres with higher efficiency than CD338 negative cells (mean ± SEM: 13.0 ± 1.080 and 1.5 ± 1.190 respectively; p=0.0004).

1.3.6.4 Mammosphere-forming efficiency of CD338^{bright}, CD338^{low} and CD338^{neg} sorted cells The use of the LSR II cytometer to analyze CD338 expression, led to the distinction of three different cell subpopulations 1) CD338$^{+/bright}$ expressing CD338 at high level and consisting just in 1% of the total cell line; 2) CD338^{neg} not expressing CD338 and representing about 20% of the total cell line; 3) CD338$^{+/low}$ expressing CD338 at an intermediate level and consisting in about 79% of the total cell line (see section 3.1.2.2 and Figure 3.5). I thus sorted these three cell subsets and performed a mammosphere formation assay. For the first generation mammospheres, I plated sorted cells at 1,000 cells/well and I counted the number of spheres in each well after 3 weeks. As shown in figure 3.12, unsorted cells formed 4.2 spheres/well ± 2 (SD). CD338^{bright} and CD338^{low} formed a higher number of spheres than unsorted cells (7.5 ± 1.8 and 6.6 ± 2.5, respectively), whereas CD338^{neg} formed a lower number (3.8 ± 1.6). So CD338^{bright} cells displayed the highest MFE, whereas CD338^{neg} displayed the lowest one. At the first generation, MFE of CD338^{bright} cells was 2 fold higher than that of CD338^{neg} . First generation mammospheres were dissociated to obtain a single cell suspension that was used to perform the second generation mammosphere assay by plating cells at a density lower than in first generation (100 cells/well instead of 1000 cells/well). At the second generation the MFE difference between CD338^{bright} and CD338^{neg} was even more pronounced (Figure 3.12). Indeed, I counted 1.5 spheres/well from CD338^{bright} and 0.5 spheres/well from CD338^{neg}. Finally, at the third generation assay, CD338^{neg} sorted cells were not able to form spheres (Figure 3.12). It's very interesting to note that, between the different sorted subsets, there was a difference not only in the number of spheres, but also in their size. At the first and the second generation, mammospheres originated by CD338^{bright} and CD338^{low} sorted cells were bigger than those originated by CD338^{neg} sorted cells (Figure 3.13).

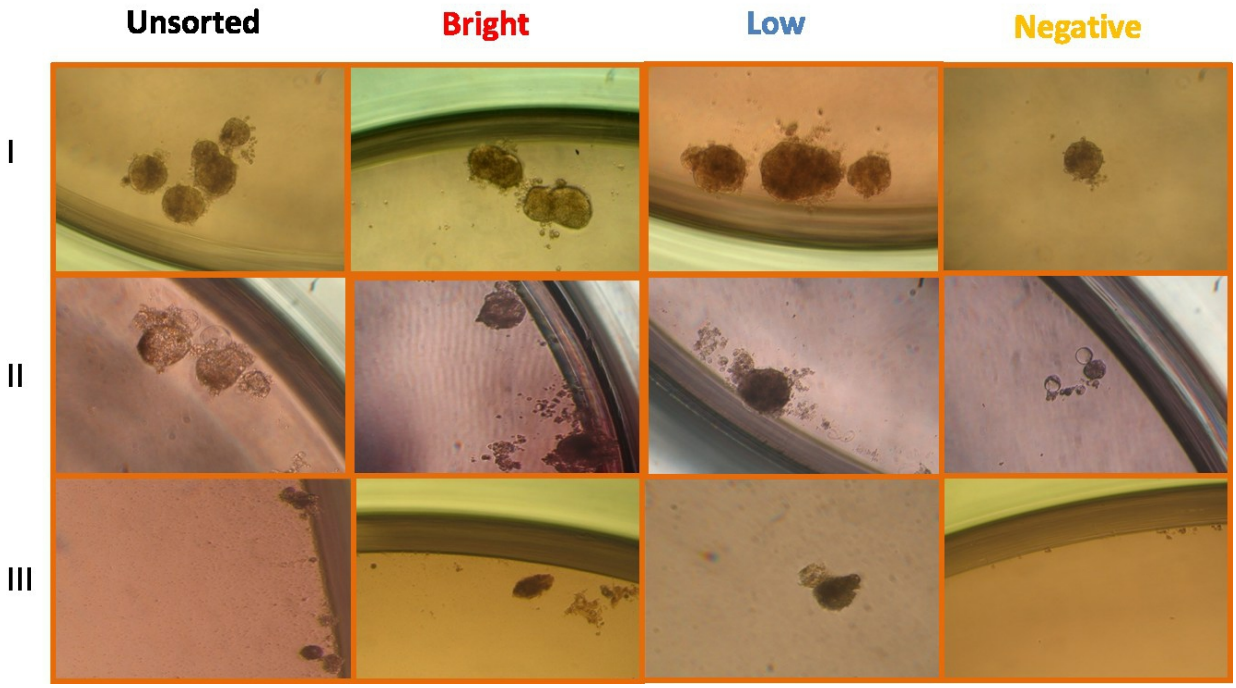


Figure 3.12. Mammosphere-forming efficiency of CD338^{bright}, CD338^{low} and CD338^{neg} sorted cells. Cells from cell line in adherent conditions were stained with anti-CD338 mAb and analyzed by flow cytometry. CD338^{bright}, CD338^{low} and CD338^{neg} cells were separated through cell sorting and plated in non-adherent conditions at a low density to generate mammospheres. The upper panels **(I)** show the result of the first generation mammospheres: MFE ± SD: 4.2 ± 2 (unsorted cells), 7.5 ± 1.4 (CD338^{bright}), 6.6 ± 2.5 (CD338^{low}) and 3.8 ± 1.6 (CD338^{neg}). The central panels **(II)** show the result of the second generation mammospheres: MFE ± SD: 1.1 ± 0.9 (unsorted cells), 1.5 ± 1.3 (CD338^{bright}), 0.9 ± 0.9 (CD338^{low}) and 0.5 ± 0.9 (CD338^{neg}). The lower panels **(III)** show the result of the third generation mammospheres: unsorted cells: 3 spheres/44wells, CD338^{bright}: 8 spheres/44wells, CD338^{low}: 5 spheres/44wells, CD338^{neg}: any spheres.

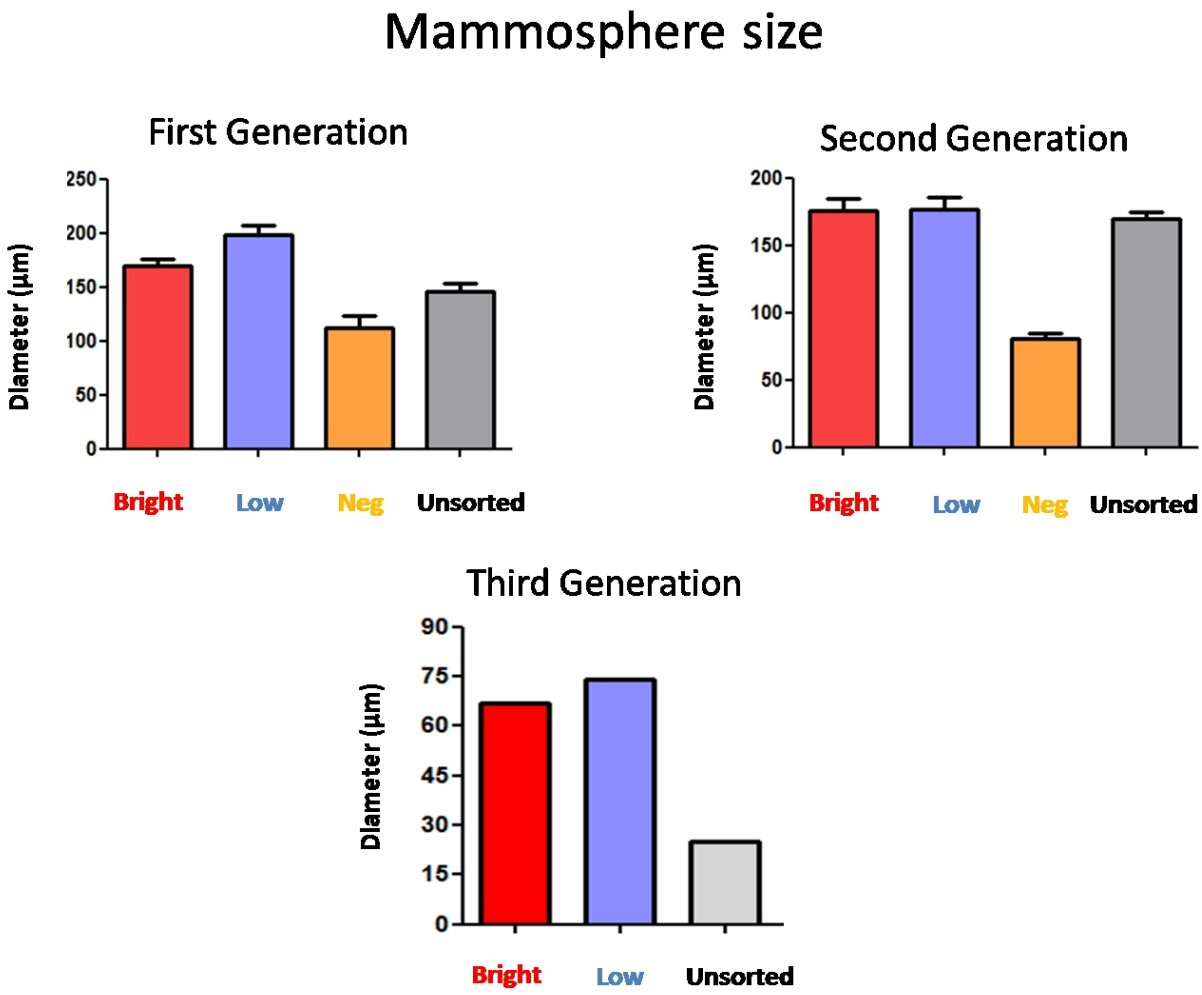


Figure 3.13. Size of mammospheres originated by CD338^{bright} CD338^{low} and CD338^{neg} sorted cells. At the first and second generation, mammospheres originated by CD338^{bright} and CD338^{low} sorted cells were bigger than those originated by CD338^{neg} sorted cells. First generation mammospheres: size (µm) ± SD: 145.8 ± 15.28 (unsorted cells), 169.9 ± 14.66 (CD338^{bright}), 197.9 ± 27.28 (CD338^{low}) and 112.5 ± 18.75 (CD338^{neg}). Second generation mammospheres: size (µm) ± SD: 170.0 ± 8.32 (unsorted cells), 176.1 ± 15.11 (CD338^{bright}), 176.5 ± 15.17 (CD338^{low}) and 80.96 ± 7.03 (CD338^{neg}). Third generation mammospheres: Any mammospheres were generated from CD338^{neg} sorted cells.

3.1.7 Soft agar colony formation assays

3.1.7.1 Soft agar colony formation assay with sorted cell subpopulations

The soft agar colony assay is a commonly accepted method to assess anchorage-independent growth and it is considered a hallmark of cancer cells. To evaluate the transformation potential of three cell populations differing in regards to their CD338 antigenic phenotype, cells were sorted out, cultured few days to allow cell recovery and plated in soft agar. As shown in figure 3.14, both ABCG2 bright and low populations gave rise to colonies, while ABCG2 negative were found to be poorly transformed. In details, $CD338^{+/bright}$ and $CD338^{+/low}$ cells formed 3450 ± 481 and 2695 ± 117 colonies respectively, while $CD338^{neg}$ formed just 948 ± 90 colonies. It is important to consider that $CD338^{neg}$ sorted population was contaminated by $CD338^{+/low}$ cells (see section 3.1.2.2), giving likely an explanation to the few colonies observed. In conclusion, CD338 positive cells are transformed in contrary to the negative ones.

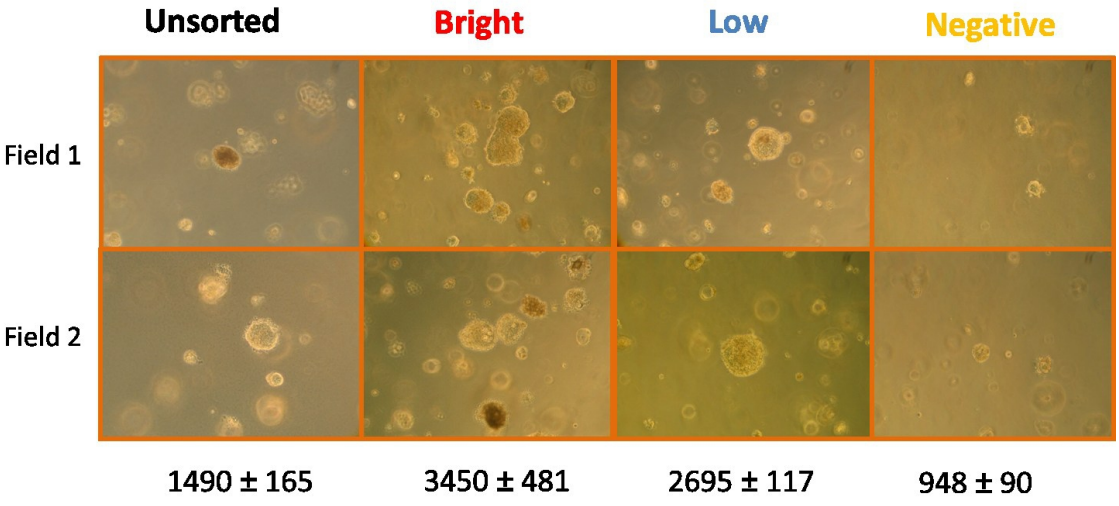


Figure 3.14. Soft agar colony formation assay with $CD338^{+/bright}$, $CD338^{+/low}$ and $CD338^{neg}$ populations. $CD338^{+/bright}$, $CD338^{+/low}$ and $CD338^{neg}$ cells were sorted out from HCC1937 cell line, put in adherent conditions for few days to allow cell recovery after cell sorting, and then used to perform the colony assay. Colonies were counted after 4 weeks. Number of colonies are indicated for 5 x 10^4 plated cells (± standard deviation, n = 3).

3.1.7.2 Soft agar colony formation assay with unsorted cells treated with cross-linker

In order to improve the cell sorting purity, I tried to use a covalent cross-linker able to stabilize the binding of 5D3 antibody to its CD338 antigen (see section 3.1.2.1). Surprisingly, sorted cells gave rise to any colonies in soft agar, suggesting that the linker actually do interfere with ABCG2 properties and that loss of ABCG2 activity avoids survival/growth in low adherent conditions (Figure 3.15). To address this question, we assessed the effect of ABCG2 depletion through RNA interference on the cell transformation potential. Among the five shRNA tested, four were successful as functionally judged by the ABCG2 down-regulation (Figure 3.16a). shRNA A was excluded for a toxicity problem. I next sought to confirm these results through assessment of ABCG2 at the protein level. Surprisingly, in all 4 knockdown cells just $CD338^{bright}$ cells disappeared, while $CD338^{low}$ cells were still present (Figure 3.16b). This data indicated that, although the silencing was efficient (78%-92%), it was not enough to completely inhibit the synthesis of ABCG2 surface protein, avoiding further investigation on the role of ABCG2 in cell transformation.

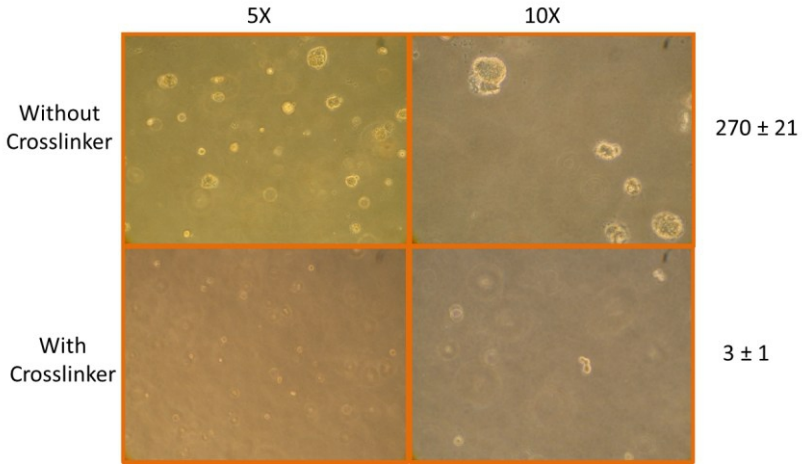


Figure 3.15. Effect of cross-linker on colony formation ability of unsorted cells. Unsorted cells were detached and divided in two aliquots. One was pre-incubated with cross-linker before CD338 staining and the other was not (see Materials & Methods). Then cells were used to perform the colony assay. Colonies were counted after 3 weeks. Number of colonies are indicated for 5 x 10^4 plated cells (± standard deviation, n = 3).

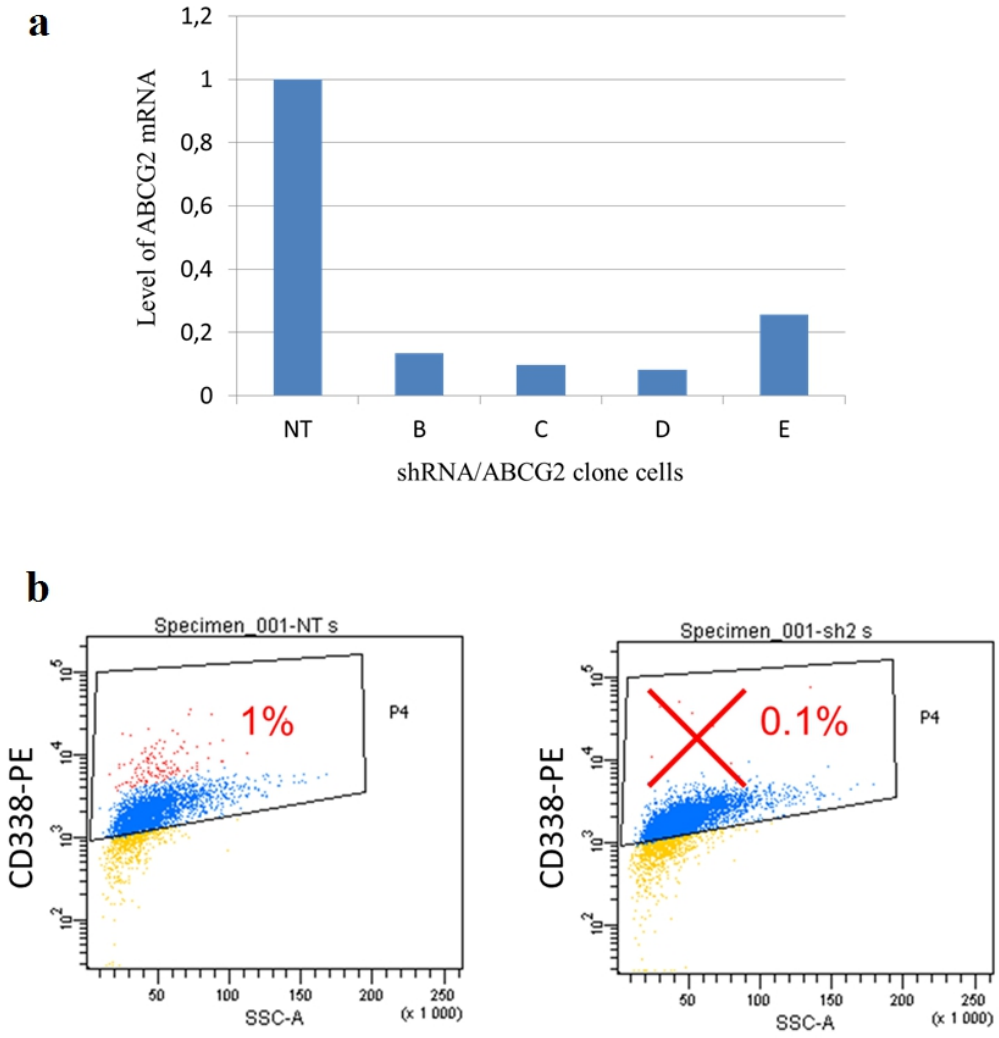


Figure 3.16. Suppression of mRNA and surface protein expression by shRNA/ABCG2. **(a)** Relative mRNA expression levels of ABCG2 in knockdown cells, as estimated by qPCR. The level of mRNA expression in the parental HCC1937 cell line was set = 1.0 for reference purposes. NT= HCC1937 cells infected with viral particles that had not been transfected with any shRNA constructs; B to E = shRNA/ABCG2 clones cells. *HPRT1* was used as a reference gene. **(b)** FACS analysis of ABCG2 surface protein expression. Just $CD338^{bright}$ cells (red events) were decreased from 1% (of the total cell line) in NT cells (left panel) to 0.1% in silenced cells (right panel).

3.1.8 Tumorigenic assays

3.1.8.1 Increase of CD338^{bright} cells in NOD/SCID mice

To evaluate the tumorigenicity of the HCC1937 cell line, I either injected subcutaneously 2×10^6 or 4×10^5 HCC1937 cells into the left and right flanks of four NOD/SCID immunocompromised mice, respectively. Tumor tissues were excised, digested to single cell suspensions and analyzed by flow cytometry (see Materials and Methods). Percentages of CD338^{bright} cells were determined in viable (SYTOX$^-$) human (HLA-ABC$^+$) epithelial (EpCAM$^+$) gated tumor-derived cells (Figure 3.17b). The CD338^{bright} subpopulation was enriched in tumor-derived cells in comparison with the parental cell line and mammosphere-derived cells: about 40% (mean ± SEM: 37.50 ± 1.079) in tumor tissues, 1.4% in adherent cell line and 9.7% in mammospheres (Figure 3.17a,c). To compare the tumorigenicity of CD24^{hi} and CD24$^-$ cells, an equivalent number of cells of both populations were injected in the left and right flanks of 5 NOD/SCID mice, respectively. Three mice were injected with $5x10^5$ cells and 2 mice with $5x10^4$ cells. There was not a significant difference in the tumorigenicity of the two CD24 populations but, interestingly, tumor tissues originated from CD24^{hi} cells contained a significantly higher percentage of CD338^{bright} cells than tumor tissues originated from CD24$^-$ cells (mean ± SEM: 55.25 ± 3.425 and 40.50 ± 0.500, respectively; $p<0.05$; Figure 3.18). This data confirm the correlation between CD24 and CD338 in HCC1937 cell line (see section 3.1.1.1 and figure 3.2). To compare the tumorigenicity of CD388^{bright} and CD338^{neg} cells, equal numbers of the two sorted cell subsets were similarly injected in the left and right flanks of 2 NOD/SCID mice, respectively. Consistently, a difference in tumor growth was observed between the two populations, the CD388 negative being indolent or giving rise to slow proliferating tumor. Although the number of animals is too low to consider these results as significant, these data suggested that CD388^{bright} cells are more tumorigenic than CD338^{neg} cells. I repeated these xenograft experiments by sorting the three

$CD338^{bright}$ $CD338^{low}$ and $CD338^{neg}$ subpopulations and injecting them in nude mice (see section 3.1.8.2). Interestingly, tumor tissues originated from $CD388^{bright}$ cells in NOD/SCID mice contained the highest percentage of $CD338^{bright}$ cells between all tumors originated in NOD/SCID mice (i.e. tumors from total cell line and tumors from $CD24^{hi}$ and $CD24^{neg}$ sorted cells), whereas tumor tissues originated from $CD388^{neg}$ cells contained the lowest one (90% and 35%, respectively; Figure 3.18).

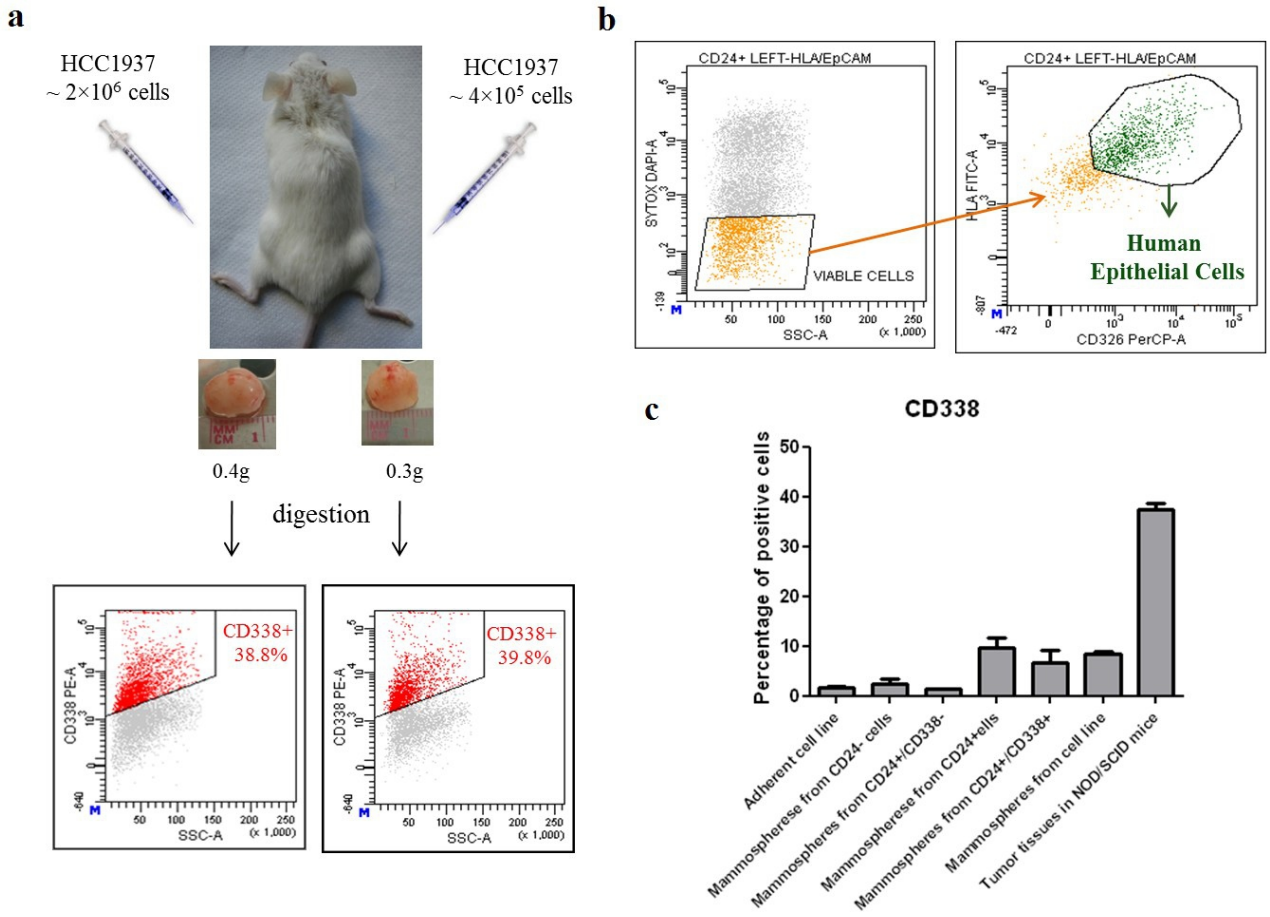


Figure 3.17. Expression of CD338 in tumors originated from unsorted HCC1937 in NOD/SCID mice. (a) Two different numbers of HCC1937 cells (2×10^6 and 4×10^5) were injected subcutaneously into the left and right flanks of NOD/SCID mice, respectively. Dot plots show the percentages of $CD338^{bright}$ cells in the cell suspensions obtained from digestion of tumor tissues. **(b)** Gating strategy used to analyze by flow cytometry the expression of CD338 in xenografts. Percentages of $CD338^{bright}$ cell subpopulation were determined in viable ($SYTOX^-$), human ($HLA\text{-}ABC^+$), epithelial ($EpCAM^+$) gated cells. **(c)** The histogram shows the percentage of $CD338^{bright}$ cells in adherent cell line, mammospheres from unsorted cell line, mammospheres from $CD24^-$, $CD24^{hi}$, $CD24^{hi}/CD338^-$, $CD24^{hi}/CD338^+$ sorted cells, and in tumor tissues from unsorted cell line.

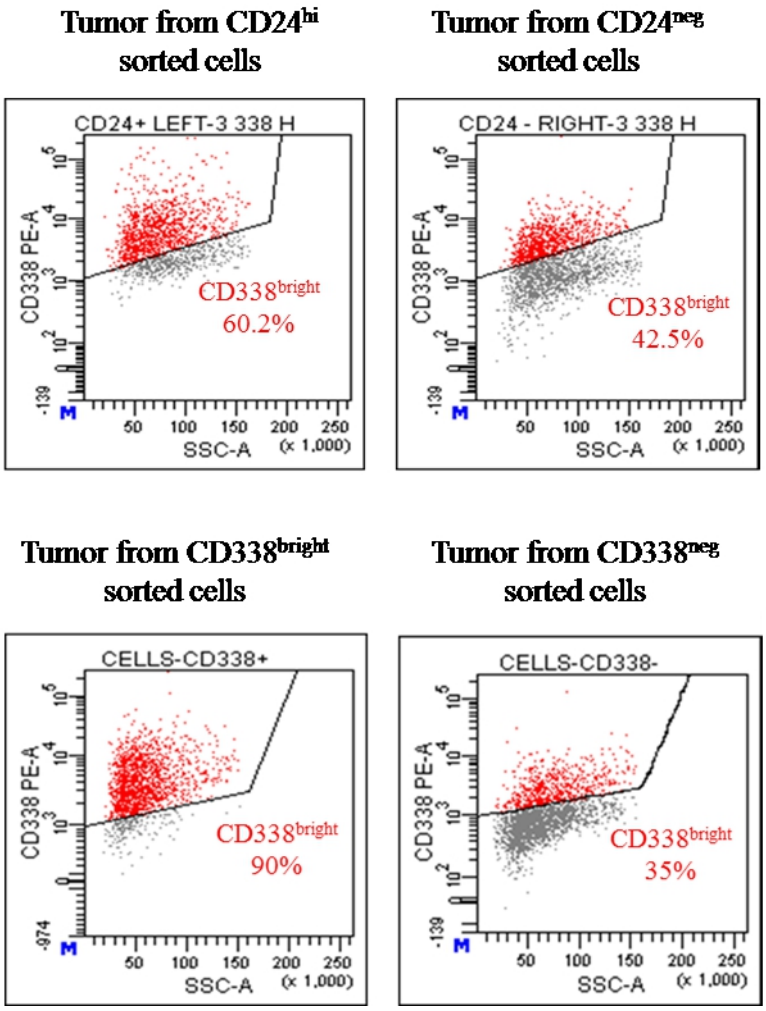


Figure 3.18. Expression of CD338 in tumors originated in NOD/SCID mice from sorted populations. Upper panels: flow cytometry analysis of tumor tissues originated from $CD24^{hi}$ and $CD24^{-}$ sorted cell populations. Lower panels: flow cytometry analysis of tumor tissues originated from $CD338^{bright}$ and $CD338^{-}$. Percentages of $CD338^{bright}$ cell subpopulation were determined in viable ($SYTOX^{-}$), human ($HLA\text{-}ABC^{+}$), epithelial ($EpCAM^{+}$) gated cells. Tumor tissues originated from $CD388^{bright}$ cells contained the highest percentage of $CD338^{bright}$ cells (90%), whereas tumor tissues originated from $CD388^{neg}$ cells contained the lowest one (35%).

3.1.8.2 Injection the three CD338 sorted subpopulations in nude mice

Results of $CD388^{bright}$ and $CD338^{neg}$ sorted cell injections in NOD/SCID mice, suggested that $CD388^{bright}$ cells are more tumorigenic than $CD338^{neg}$ cells. These data were not statistically significant due to the low number of injected mice. I repeated the xenograft experiments by sorting the three $CD338^{bright}$ $CD338^{low}$ and $CD338^{neg}$ subpopulations and injecting them in fad pats of *nude* mice. 12 mice, 4 for each of the three populations, were injected with sorted populations (10^4 cells in each mouse). $CD338^{low}$ sorted cells were the most tumorigenic (4/4 for $CD338^{low}$ *versus* 2/4 for $CD338^{bright}$ and 2/4 for $CD338^{neg}$, Table 4). As a control, unsorted HCC1937 cells were injected ($2x10^6$). Surprinsigly, most of the tumor-derived cells displayed an antigenic phenotype of luminal progenitor, i.e. $EpCAM^+/CD49f^+$ (Figure 3.19) and a $CD338^{low}$ antigenic phenotype. This result is different from that previously obtained from injection of unsorted cells in NOD/SCID mice, where tumor-derived cells displayed a heterogeneous CD338 antigenic phenotype (Figure 3.17, panel a).

Table 4. Tumorigenicity of CD338 bright, low and negative sorted populations in *nude* mice

Polulation	N of cells injected	Tumor formation efficiency*
LOW	$1x10^4$	4/4
NEGATIVE	$1x10^4$	2/4
BRIGHT	$1x10^4$	2/4

*** Number of tumors after 105-days inoculation/total mice injected**

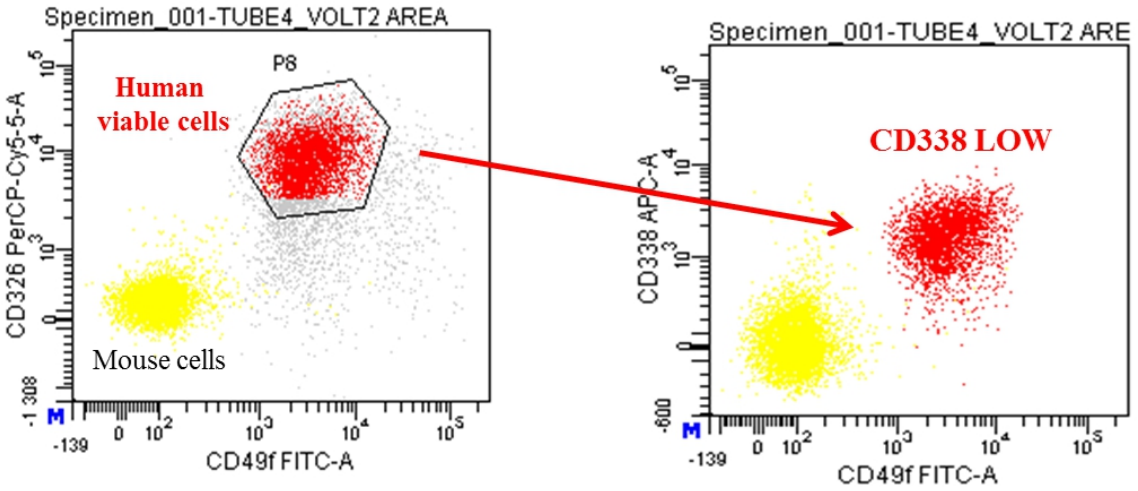


Figure 3.19. Flow cytometry analysis of tumor originated from unsorted HCC1937 cells. Eight weeks after the injection, tumor was excised, digested to a single cell suspension and analyzed by flow cytometry. Tumor contained 40% of viable cells ($DAPI^-$). 80% of viable cells were positive for the human marker HLA-ABC (red events), whereas 20% were negative (yellow events). Most human cells displayed an antigenic phenotype of luminal progenitor ($EpCAM^+/CD49f^+$, left panel). Almost all human cells were $CD338^{low}$ (right panel).

3.1.9 Expression analysis of stemness genes, EMT inducers and epithelial/mesenchymal markers in CD338 sorted cells

In breast cancer the epithelial-mesenchymal transition is highly related to tumor-initiating cell phenotype and basal subtype. So, I decided to explore the expression of stemness genes, EMT inducers and epithelial/mesenchymal markers in $CD338^{bright}$ and $CD338^{neg}$ sorted cell subpopulations. I chose as epithelial/mesenchymal markers *E-cadherin* and *vimentin* respectively, as stemness genes *Sox2, Sox9, Oct4* and *FUT4*, and as EMT inducers *Slug, Zeb1, Zeb2, Twist1*. As shown in figure 3.20, results of RT-qPCR analysis of these genes were not significant. Particularly, some genes were differentially expressed in comparison to the total unsorted HCC1937 cell line, used as control, but the differences were not strong. For example, *Sox2* and *Sox9* displayed the highest expression in $CD338^{bright}$ sorted cells, but their expression level was not strongly different to that of HCC1937 total cell line. The highest gene expression increase in $CD338^{bright}$ sorted cells was that of *Zeb2* (2.8), but it does not seem significant because there is an increase, even if less strong, also in $CD338^{neg}$ sorted cells. Surprisingly, *Oct4* displayed the highest expression in $CD338^{neg}$ sorted cells.

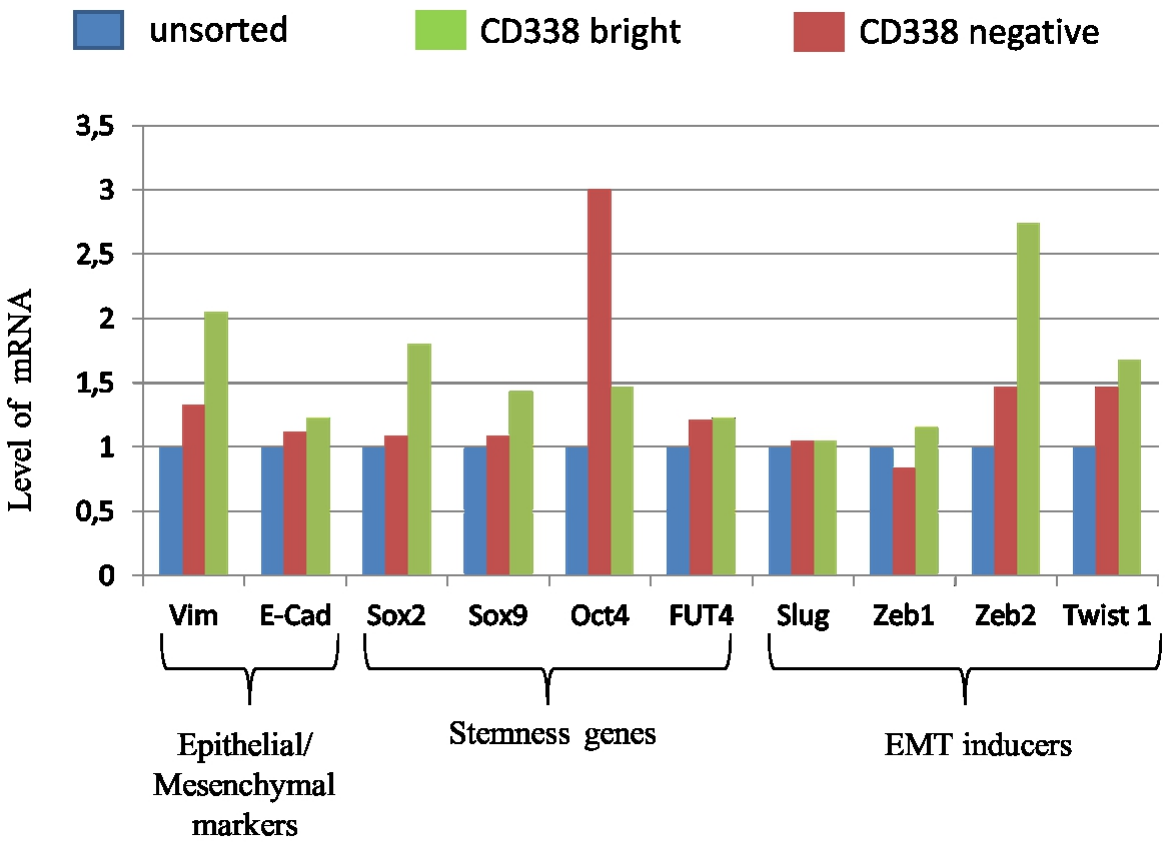


Figure 3.20. Gene expression analysis of CD338 sorted and silenced cells. Relative mRNA expression levels of epithelial/mesenchymal markers, stemness genes and EMT inducers in CD338 bright and negative sorted populations, as estimated by qPCR. The level of mRNA expression in the parental cell line (unsorted) was set = 1.0 for reference purposes. *HPRT1* was used as a reference gene.

3.2 DISCUSSION

The CSC model suggests that cancers arise from cells that can be defined, by analogy with totipotent stem cells, as cancer stem cells (CSCs) displaying self-renewal and differentiation potentials, to give rise to the phenotypically diverse cancer cell populations. However, the origin of CSCs is currently elusive. They could either originate from the malignant transformation of a stem/progenitor cell through the deregulation of the normally tightly regulated self-renewal program, or from transformation of committed cells through dedifferentiation of mature cells that reacquire some stem cell-like features including a self-renewal potential.

Breast cancer is the first human carcinoma for which a putative cancer stem cell subpopulation (BCSCs, breast cancer stem cells) has been isolated at the basis of its $CD44^{+}/CD24^{-/low}$ antigenic phenotype (Al-Hajj *et al.,* 2003). The high heterogeneity of breast cancers fuels the actual debate on the cell at origin of the different subtypes of tumors. While a hierarchical model has been proposed, suggesting that luminal, basal and claudin-low subtypes arise from luminal progenitors, basal progenitors and stem cells respectively (Prat & Perou, 2009; Prat *et al*., 2010; Perou 2011), a flow of recent observations suggests that it might be more complex than previously thought. The identification of cells at the origin of BRCA1-tumors constitutes a first, and likely not unique, example of such a complexity. The most frequent tumors arising in *BRCA1* germ-line mutation carriers have morphological features similar to those described in basal-like cancers. They indeed display a triple negative and basal-like phenotype as defined by immunohistochemistry or expression arrays (Foulkes *et al*., 2004; Lakhani *et al*., 2005; Sorlie *et al*., 2003). Initial pathological and histological studies of *BRCA1* mutant breast tumors thus suggested that these cancers may arise from basal/myoepithelial cells (Liu *et al.,* 2008). However, unexpectedly, recent studies rather

demonstrate that they originate from luminal progenitors (Lim *et al* 2009; Prat *et al*., 2010; Molyneux *et al*., 2010; Proia *et al*., 2011). Furthermore, neoplastic transformation often associates with a cell dedifferentiation process, as exemplified by the development of claudin-low tumors upon aberrant expression of Ras and Twist in luminal committed cells (Morel *et al*., 2012). To ascertain the cell at origin of the various breast cancer subtypes, a first need is undoubtedly to affine the antigenic phenotype of the different cell subpopulations including the breast cancer stem cells. The $ESA^{+}CD44^{+}CD24^{-}$ antigenic phenotype, originally identified by Clarke's group, has been shown to be restricted to tumor initiating cells of certain breast cancers (Honeth *et al*., 2008). According to this finding, Meyer and colleagues showed that both $CD44^{+}CD24^{-}$ and $CD44^{+}CD24^{+}$ cell populations sorted from ER (Estrogen Receptor) negative breast tumors are tumorigenic in murine xenograft models, indicating that xenograft initiating cells are not always restricted to the $CD44^{+}CD24^{-}$ cell population (Meyer *et al*., 2010). Although $CD44^{+}CD24^{-}$ tumorigenic cells may not display stemness features, these findings could also reveal the existence of CSCs displaying an antigenic phenotype different of the reported $CD44^{+}CD24^{-}$ phenotype. These studies, among others, thus question the relevance of CD44 and CD24 in identifying BCSCs. Obviously, there is an urgent need of additional markers to further discriminate the BCSCs.

To identify novel BCSC markers, I focused my work on basal-like breast tumors because of two types of evidences: 1) these tumors, particularly the *BRCA1* mutated basal-like breast cancers, have been proposed to arise from luminal progenitors (Lim *et al*., 2009; Molyneux *et al*., 2010; Lindeman and Visvader 2010; Perou, 2011); 2) Basal-like breast cancers do not homogeneously contain a $CD44^{+}/CD24^{-/low}$ cell population (Honeth *et al.,* 2008; Meyer *et al*., 2010). These and other studies led me to consider basal-like breast cancers, and the corresponding cell lines, as appropriate targets to identify CSCs with a phenotype different from the classical $CD44^{+}/CD24^{-/low}$ phenotype.

I particularly focused my attention on the basal-A *BRCA1* mutated cell line HCC1937, established from a primary non-metastatic IDC (Invasive Ductal Carcinoma), as these cells were previously reported to include a CSC population that do not display a $CD44^{+}/CD24^{-/low}$ antigenic phenotype (Hwang-Verslues *et al*., 2009). In support of this view, by performing a wide flow cytometry characterization of several breast cancer cell lines, I found that the surface marker CD338/ABCG2 is heterogeneously expressed and displayed a higher expression level in $CD24^{+}$ than in $CD24^{-}$ cell subpopulation of the HCC1937 cell line. Furthemore, I found an almost complete overlap between side population and CD338 overexpressing. Given that CD338/ABCG2 has been reported to be highly expressed in stem/progenitor cells and that the side population phenotype is known to be associated to stem/progenitor cells (Zhou *et al*., 2001; Cai *et al*., 2004), I thus reasoned that HCC1937 cell line could contain a putative $CD24^{+}CD338^{+}$ CSC subpopulation. This would confirm the presence, found by Hwang-Verslues, in the HCC1937 cell line of CSCs that do not harbor a $CD44^{+}/CD24^{-/low}$ antigenic phenotype.

Mammosphere formation assays confirmed the presence of cancer cells with stem cell-like properties in the CD24+ population, and enrichment in CD338 in these cells. The enrichment in CD338 overexpressing cells in mammospheres and the overlap between SP and CD338 overexpressing cells are coherent with the enrichment in SP cells in mammopsheres previously reported by Dontu (Dontu *et al*., 2003). The positive correlation between CD338 and CD24 was further confirmed by the observation that, among the CD24 positive cells, those overexpressing the stem cell marker CD338 were able to form mammospheres with higher efficiency than CD338 negative cells.

Although ABCG2 has been reported to be the major mediator of side population, on the other hand, it should be noted that only a fraction of the side population cells expresses ABCG2 (Watanabe *et al*., 2004). Both side populations and stem/progenitor cells express additional

ABC transporters (Lechner *et al*., 2002; Hirschmann-Jax *et al*., 2004). This could provide a first explanation of the partial overlap between side population and CD338 overexpressing cells in HCC1937 cell line.

Taken together, flow-cytometry analysis, examination of side population and mammosphere formation assays support the conclusion that CD338 might constitute a novel antigen, likely more reliable than CD24, to isolate BCSCs at least in such a cell line.

Unfortunately, use of this antigen has revealed some unexpected difficulties. Indeed, flow-cytometry performed with the monoclonal anti-CD338 antibody 5D3, revealed a gradient of expression of cell surface antigen rather than a negative and positive cell subpopulations. This consequently complicates the isolation of the cell populations and reduces their purity. The purity of a cell sorting is even more affected when sorted cells are rare, e.g. cancer stem cell populations, technically requiring to sort cells during long periods of time, and further complicated when the binding of the antibody to the surface antigen is not stable. As the $CD338^{bright}$ cell population constitutes 1% of the HCC1937 cells, these technical difficulties were obvious. Furthermore, 5D3 binding to its extracellular epitope strongly depends on the conformation of ABCG2 (Ozvegy-Laczka *et al*., 2005), making the binding particularly unstable. Attempts to stabilize the antibody-antigen interaction with a protein cross-linker (PMPI, p-Maleimidophenyl isocyanate) were successful, significantly increasing the purity of $CD338^{bright}$ cell sorted subpopulation (from 50-70% to 90-95%). Unfortunately, the crosslink of the transporter turned out to down-modulate its activity as demonstrated by the lack of colonies when unsorted cells were plated on soft agar. Although disappointing, these experiments likely highlight the key role of the transporter for the transformation potential of the cells, a hypothesis that we failed to address through RNA interference due to the limited efficiency of the shRNAs employed.

Nonetheless, taking benefit of LSR II, a third cell subpopulation expressing an intermediate level of CD338 has been unveiled. HCC1937 cells are thus constituted of three populations 1) $CD338^{+/bright}$ expressing CD338 at high level and consisting just in 1% of the total cell line; 2) $CD338^{neg}$ not expressing CD338 and representing about 20% of the total cell line; 3) $CD338^{+/low}$ expressing CD338 at an intermediate level and consisting in about 79% of the total cell line. Despite the fact that this cytometer is not an analyzer, thanks to it, the three cell populations could be sorted out with a FACSAria cell sorter to assess their functional properties.

Colony soft-agar assays demonstrated that CD338 positive cells are the only subpopulations displaying a transformation potential. Hwang-Verslues *et al*. found that $CD24^{+}$ cells of HCC1937 cell line form more colonies than $CD24^{-}$ cells in soft agar assay. As we observed that CD338 positive cells also overexpress CD24, we conclude from the soft agar colony assays that $CD24^{+}CD338^{+}$ cells constitutes the only subpopulation of the HCC1937 cell line displaying a transformation potential. When xenografted in NOD/SCID and *nude* mice, CD338 negative were confirmed to be indolent while the CD338 positive were tumorigenic. This observations strengthen our conclusion that ABCG2 specifically map CSCs. To date, numerous studies have been conducted to analyze (i) ABCG2 expression in stem cells and human tumors and (ii) its activity as transporter in relationship to the SP phenotype. Conversely, there are still very few studies assessing the tumorigenic potential of ABCG2+ cells in comparison to ABCG2- ones. Patrawala and colleagues found that, although ABCG2 preferentially mark proliferating cells, ABCG2 positive cells are not more tumorigenic than the corresponding ABCG2 negative cells (Patrawala *et al*., 2005). The conflict of our results with those of Patrawala could be explained by the different models used in the two studies. Patrawala and colleagues used cell lines from different types of tumors and they chose MDA-MB-435 as breast cancer cell line. MDA-MB-435 is a basal-B cell line, so it belongs to a

molecular subtype different from that of HCC1937 cell line, i.e. basal-A. The difference of our results with those of Patrawala *et al*. could also be explained by the fact that a single marker is not sufficient to identify stem cells and the combination of CD24 and CD338 is likely to be more stringent. This is also supported by the observation that tumorigenic assays performed with $CD24^{hi}$ and $CD24^{-}$ sorted cells did not reveal any significant differences in the tumorigenicity of the two CD24 populations. Clearly, CD24 cannot be used on its one to identify stem cells, whereas CD338 in this regard is more relevant and the combination of the two antigens is likely to provide a better segregation. Our results also suggest a relationship between stem-like and tumorigenic properties of HCC1937 cells, as long as similar results have been obtained from mammosphere and xenograft assays. Actually, mammosphere assays revealed that, among $CD24^{+}$ cells, those overexpressing CD338 displayed a higher mammosphere forming efficiency, and xenograft assays revealed a high tumorigenic potential for CD338 positive cells in comparison to CD338 negative ones, but not for CD24 positive cells in comparison to CD24 negative ones. Nonethelesss, a positive correlation between CD24 and CD338 was confirmed by the increase of $CD338^{bright}$ in tumor tissues originated from $CD24^{hi}$. The relevance of CD338 to isolate stem-like/tumorigenic cells from HCC1937 cell line, was also confirmed by tumorigenic assays performed with the unsorted cell line, as $CD338^{bright}$ subpopulation was strongly enriched in HCC1937-derived tumors in comparison with the parental cell line.

Asymmetric division is one of the main properties of stem cells (Boman *et al*., 2007; Morrison and Kimble 2006). The analysis of $CD338^{bright}$ and $CD338^{low}$ sorted sub-populations, after maintaining them in culture for several weeks, revealed that the antigenic phenotype of $CD388^{low}$ cells remained stable and homogeneous, whereas $CD338^{bright}$ cells gave rise to bright and low cells suggesting a filiation from $CD338^{bright}$ cells to $CD338^{low}$ cells. Our results are in agreement with the findings that, in several tumor cell lines, in just ~1% of

the ABCG2+ dividing cells, ABCG2 segregated asymmetrically to mainly one daughter cell (Patrawala *et al.*, 2005). This observation suggests that just some ABCG2+ expressing cancer cells (1%, so a very few cells) might be undergoing asymmetrical cell division. Moreover, it has been recently reported that ABCG2+ cells purified from human hepatocellular carcinoma cell lines showed evidence for self-renewal, generating both ABCG2+ and ABCG2− progenies during subculture, and a higher proliferative activity (Zen *et al*. 2007). Furthermore, ABCG2 is known to be down-regulated during hematopoietic stem cell differentiation and is expressed at a low level in mature cells compared with progenitor cells (Zhou et al. 2001; Scharenberg *et al*. 2002). The same phenomenon also occurs in human neural, retinal and embryonic stem cells (Islam *et al*. 2005; Bhattacharya *et al*. 2007; Apati *et al*. 2008). The highly regulated expression of ABCG2 suggests that ABCG2 may play a regulatory role in maintaining stem cells in an undifferentiated state. Considering that ABCG2 acts as an efflux pump, it is also possible that expression of ABCG2 in human progenitor cells critically expels substrates that are necessary for lineage differentiation, thereby blocking stem cell differentiation. We also found that $CD338^{low}$ cells in tumors originated from unsorted HCC1937 cells display an antigenic phenotype of luminal progenitors, i.e. EpCAM+/CD49f+. This result, together with the observation that $CD338^{bright}$ cells generate $CD338^{bright}$ and $CD338^{low}$ cells in subcultures, strongly supports the hypothesis that $CD338^{low}$ cells could constitute the progeny of the most immature $CD338^{bright}$ cells, and in particular, the luminal progenitor cells likely blocked at this stage of development. This hypothesis would also explain the fact that $CD338^{low}$ cells constitute the most part of HCC1937 cell line (79%). The weak purity of CD388 negatively sorted out cell population avoided us performing filiation experiments with this subpopulation; thus, we can just hypothesize that $CD338^{neg}$ cells could constitute the progeny of $CD338^{low}$ population and so the most differentiated cells, as also suggested also by their low mammosphere-forming efficiency. Figure 3.21 shows a possible

model of filiation between CD338^{bright}, CD338^{low} and CD338^{neg} cell subpopulations in HCC1937 cell line. This model would be in agreement with that proposed by Lim *et al*. who found an increased number of luminal progenitors in breast tissue from BRCA1 mutation carriers. Indeed, HCC1937 is a cell line established from a BRCA1 mutated basal-like breast cancer.

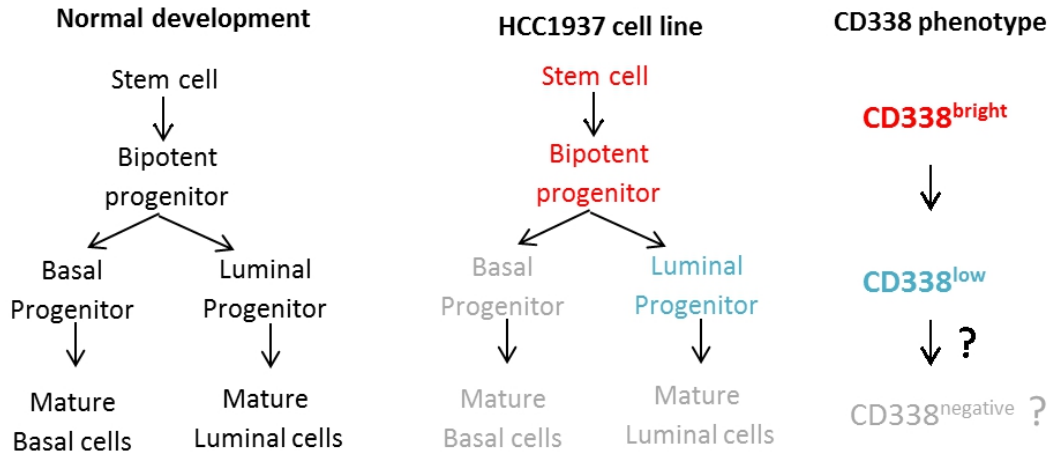


Figure 3.21. Schematic representation of the possible filiation between the different CD338 cell subsets in the HCC1937 cell line

Obviously, to conclude that CD338 might constitute a novel antigen reliable to isolate BCSCs from this subtype of breast cancers, our results should be extended to other BRCA1 mutated basal-like cell lines and then tested directly on human breast cancer tissues. Actually, the accumulation of $CD338^{low}$ cells in HCC1937 cell line could also be the result of *in vitro* selection. In that case, the fact that after several weeks in culture $CD338^{bright}$ sorted cell population contain also $CD338^{low}$ cells, could be explained through an *in vitro* equilibrium between the two populations. Anyway, the different possibilities about the nature of $CD338^{low}$ cells in the HCC1937 cell line, could be elucidated by their *in vivo* behavior and so by the analysis of tumors originated from the sorted cell subpopulation. Indeed, if tumors originated from $CD338^{low}$ sorted cells would contain only $CD338^{low}$ cells, the first hypothesis that $CD338^{low}$ cells would constitute the luminal progenitor cells likely blocked at this stage of development, could be confirmed. The model proposed by Lim *et al*. would then be extrapolated to the HCC1937 cell line. Another possibility is that tumors originated from $CD338^{low}$ sorted cells contain both $CD338^{low}$ and $CD338^{bright}$ cells, this could indicate the ability of $CD338^{low}$ cells to dedifferentiate *in vivo* thus generating the more immature $CD338^{bright}$ cells. Ongoing experiments would help us to discriminate between these two possibilities.

3.3 MATERIALS AND METHODS

3.3.1 Cell Culture

The human breast cancer continuous cell line HCC1937, was purchased from the American Type Culture Collection (ATCC, Rockville, MD, USA). The cells were initially cultured in RPMI (Sigma-Aldrich Corporation, St. Louis, Missouri, USA), 10% fetal bovine serum (FBS; Hyclone, Logan, UT, USA) and 1% Ultraglutamine 1 (Lonza Verviers, Belgium). Later, due to their high metabolism, cells were cultured in IMDM (Invitrogen), 20% FBS (Gibco).

To obtain the high numbers necessary for cell sorting cells were grown in 15 cm dishes in a humidified atmosphere of air containing 5% CO_2 until confluence, then they were detached by trypsin/EDTA (Sigma-Aldrich Corporation, St. Louis, Missouri, USA).

To follow the expression of CD338 after cell sorting, cell subpopulations sorted out from HCC1937 cell line were seeded and grown in 12-well, 6-well or 10 cm dishes. The human breast cancer continuous cell lines MCF7, Hs578T and BT-20 were cultured in DMEM, 10% FBS. During the study, all cell lines were periodically tested for the absence of bacteria, fungi and mycoplasma contaminations.

3.3.2 Flow cytometry analysis and cell sorting

3.3.2.1 Expression analysis of surface markers in breast cancer cell lines

Antigens and antibodies

On breast cancer cell lines mentioned above, MCF-7, HCC1937, BT-20 and Hs578, the expression of 30 surface antigens was evaluated: CD9 (p24), CD10 (CALLA), CD24 (HSA), CD26 (DPPIV), CD29 (β1 integrin), CD44 (H-CAM), CD47 (Rh-associated protein), CD49b (α2 integrin), CD49f (α6 integrin), CD54 (ICAM), CD55 (DAF), CD59 (MIRL), CD61 (β3 integrin), CD66b (CEACAM8), CD66c (CEACAM6), CD81 (TAPA1), CD90 (Thy1), CD105 (Endoglin), CD133 (prominin1), CD151 (PETA-3), CD164 (MUC-24), CD165 (AD2), CD166 (ALCAM), CD184 (CXCR4), CD200 (OX2), CD227 (MUC1), CD324 (E-Cadherin), CD326 (EpCAM), CD338 (ABCG2), CD340 (Her2). These surface antigens include molecules involved in processes such as cell adhesion, migration, apoptosis, metastasis, cell signalling, proliferation, differentiation and molecules known to be stem cell and/or cancer stem cell markers (Table 2, Results section).

Multi-color flow cytometry was performed by using anti-human monoclonal antibodies (MoAbs) that were conjugated with phycoerythrin (PE), fluorescein isothiocyanate (FITC), phycoerythin-Cy7 (PE-Cy7) or Alexa Fluor 647. PE-conjugated MoAbs against CD10, CD29, CD54, CD55, CD59, CD61, CD151, CD166, CD200, CD340 and FITC-conjugated MoAbs against CD9, CD26, CD47, CD49b, CD49f , CD66b, CD66c, CD81, CD90, CD164, CD165, CD227, CD324 and CD326 were from BD Biosciences and BD Pharmingen, (San Jose, CA); PE-coniugated MoAb against CD133 from Miltenyi Biotech (Auburn, CA, USA); AlexaFluor647-conjugated MoAb against CD24, PE-Cy7-conjugated MoAb against CD44 and PE conjugated MoAb against CD338 from Biolegend (San Diego, CA, USA); PE-conjugated MoAb against CD184 from Immunotech (Marseille, France); PE-conjugated MoAb anti-CD105 from Serotec (Kidlington, Oxford, UK).

Sample preparation

For expression analysis of the 30 surface markers a four-color flow cytometric strategy was performed. In particular, cells were stained with MoAbs anti-CD24-AlexaFluor 647 and anti-CD44-PECy7 combined with pairs of antibodies, conjugated with other two different fluorochromes, against surface antigens explored as potential novel BCSC markers (Table 5). Thus, each experiment consisted of 14 tubes, each containing MoAb anti-CD24-AlexaFluor647, MoAb anti-CD44-PE-Cy7, one PE-conjugated and one FITC-conjugated antibody. In all expression analysis experiments, an analysis buffer (AB) was used to prepare cells: RPMI without red phenol (Invitrogen), 1-2% FBS (Gibco), 10U/ml DNase (Sigma-Alderich). After enzymatic detachment from saturated cultures, cells were counted, resuspended in the analysis buffer at $5x10^5$ for 100µl and stained by incubation at 4°C for 20 minutes with the appropriate amount of above mentioned MoAbs. After staining, all samples were washed twice with the analysis buffer, centrifuged and resuspended in 0.5 ml of FACS buffer (FACS Flow Sheat Fluid, BD Biosciences) for FACS analysis. A few minutes before FACS acquisition cells were incubated at RT in the dark with a vital dye (SytoxBlue or DAPI, Invitrogen), to exclude dead cells from the analysis. In all experiments, a negative control was prepared to exclude the signal background caused by the cellular autofluorescence. The negative control consisted in HCC1937 cells treated with the same procedure but except the antibody staining. (see 6.2 of Flow Cytometry Appendix).

Table 5. Four-color flow cytometry panel for expression analysis of surface markers in breast cancer cell lines

Tube	FITC	PE	PE/Cy7	AlexaFluor647
1	CD49f	CD338	CD44	CD24
2	CD90	CD133	CD44	CD24
3	CD47	CD200	CD44	CD24
4	CD227	CD10	CD44	CD24
5	CD324	CD29	CD44	CD24
6	CD66c	CD61	CD44	CD24
7	CD49b	CD184	CD44	CD24
8	CD164	CD340	CD44	CD24
9	CD9	CD54	CD44	CD24
10	CD26	CD55	CD44	CD24
11	CD81	CD151	CD44	CD24
12	CD66b	CD59	CD44	CD24
13	CD165	CD166	CD44	CD24
14	CD326	CD105	CD44	CD24

Cytometers

The samples were analyzed by using FACSAria I flow cytometer (Becton Dickinson, Franklin Lakes, NJ). FITC, PE and PE-Cy7 fluorescences were determined by a 488nm excitation line and detected by 530/30nm, 585/42nm and 780/60nm filters, respectively. AlexaFluor-647 fluorescence was determined by a 633nm excitation line and detected by a 660/20nm filter (see sections 6.1 and 6.4 of Flow Cytometry Appendix). For each sample run, 10,000 to 20,000 events were recorded and analyzed. In the initial phase of the study, I could use only the 488nm laser of FACSAria I flow cytometer to excite PE which was the fluorochrome conjugated to the anti-CD338 MoAb. This laser allows to discriminate only CD338 positive and negative cells. Later, I had the opportunity to analyze the expression of CD338 with the 561nm laser of BD LSR II four-laser flow cytometer (see section 6.4 of Flow Cytometry Appendix). This laser is able to better discriminate the different PE fluorescences, so it allowed to distinguish, among the CD338 negative cells, the real negative cells from a subpopulation of cells expressing CD338 at low level, named $CD338^{+/low}$.

Analysis strategy of cytometric data

The samples were analyzed by using the FACSDiva software (Becton Dickinson, Franklin Lakes, NJ). To define the target cell population to analyze for the expression of 28 surface markers, we used a three gating strategy: first, to exclude dead cells and debris, cells were gated on a two physical parameters dot plot measuring forward scatter (FSC) vs side scatter (SSC) (see sections 6.1 and 6.2 of Flow Cytometry Appendix). Doublets were excluded by gating cells on FSC-Height vs FSC-Area dot plots (see section 6.2 Flow Cytometry Appendix) and finally, also to exclude dead cells, we gated Sytox Blue or DAPI negative cells. The expression level of each surface marker in the different cell lines was reported as percentage of positive cells in Count *vs* FITC- or PE-CD histograms. Furthermore, I also defined the expression level of each antigen in the different CD44/CD24 subpopulations, since each FITC-CD or PE-CD antibody was combined with anti-CD44 and anti-CD24 antibodies.

3.3.2.2 Cell sorting of CD24 and/or CD338 cell subpopulations of HCC1937 cell line

Cell expansion for cell sorting

For each cell sorting, HCC1937 cell line was expanded until enough cells were available to obtain the appropriate number of sorted cells necessary for the downstream experiments. During the whole study, the limiting subpopulation consisted in cells expressing CD338 at high level ($CD338^{bright}$ cells) that represent only 1% of the total HCC1937 cell line. So, to sort the different CD338 subpopulations, cells were expanded up to $150x10^6$ in 150x25mm cell culture dishes. This level of cell expansion allowed me to obtain a number of $CD338^{bright}$ sorted cells enough to perform downstream experiments such as soft agar assays, mammosphere assays and RT-qPCR.

Sample preparation

During the preparation of cells for cell sorting, a sorting buffer (SB) was used: RPMI without red phenol (Invitrogen), 1-2% FBS (Gibco) 10U/ml DNase (Sigma-Alderich) and 2.5mM EDTA. After enzymatic detachment from saturated cultures, cells were counted, resuspended at $50x10^6$ for ml in the SB and stained by incubation with the appropriate amount of CD338 and/or CD24 at 4°C for 1h. To increase the purity of the cell sorting, before CD338 staining, cells were incubated with 1mM PMPI (p-Maleimidophenyl isocyanate) (Thermo Scientific Pierce), a cross-linker able to stabilize the epitope of CD338 antibody. After staining, samples were washed twice with the SB, and resuspended at $20x10^6$ for ml of the SB. Cells were aliquoted in 5ml polypropylene tubes by filtration with 50µm filters (Partech). Finally, cells were incubated few minutes at RT in the dark with a vital dye, SytoxBlue (Invitrogen) or DAPI (Invitrogen).

Cell sorting

Live cell sorting (see section 6.3 of Flow Cytometry Appendix) experiments were performrd using BD FACSAria I with 100µm nozzle. PE fluorescence of CD338 was determined by a 488nm excitation line and detected by 585/42nm filters, whereas AlexaFluor 647 fluorescence of CD24 was determined by a 633nm excitation line and detected by 660/20nm filter. In the initial phase of the study, just CD338 positive and negative subsets were usually sorted and further studied. Later, I performed three-ways cell sorting, that means three different subpopulations, $CD338^{+/bright}$ $CD338^{+/low}$ and $CD338^{neg}$, were sorted. Sorted cells were collected in a collecting buffer (CB) - RPMI without red phenol (Invitrogen), 20% FBS (Gibco) 10U/ml DNase (Sigma-Alderich) and 2% penicillin/streptomycin (Invitrogen) - and used to perform mammosphere assays, soft agar colony assays, RT-QPCR and *in vivo* tumoigenic experiments.

3.3.3 Side population in HCC1937 cell line

Progenitor/stem cells are able to actively extrude the fluorescent Hoechst 33342 dye, and so, they are negative for this dye. The negative-Hoechst 33342 cell population is called Side Population (SP) and ABCG2 seems to be one of the major mediators of side population. To study the side population in HCC1937 cell line, the method described by Goodell *et al*. was used. Cells ($1x10^6$ cells/ml) were incubated in pre-warmed medium (RPMI without red phenol 20% FBS). Freshly prepared Hoechst-33342 (Sigma-Alderich) at 5µg/ml was added and cells were incubated for 90 min at 37°C. In each experiment one sample set was treated with 5µM Reserpine which inhibits efflux of the Hoechst 33342 and therefore functions as a negative control. After incubation, cells were centrifuged at 2000 rpm for 5 minutes at 4°C, resuspended in ice-cold RPMI and stained with 2µg/ml Propidium Iodide (to exclude dead cells from analysis) by incubation in the dark at 4°C for 10 minutes. Finally, cells were washed once with ice-cold RPMI and analyzed by FACSAria I flow cytometer. The Hoechst dye was excited at 375 nm by the Near UV laser option and its fluorescence was measured by dual wavelength: 450/20 nm BP filter (Hoechst blue) and 675 EFLP optical filter (Hoechst red, see section 6.1 of Flow Cytometry Appendix). The side population was gated as defined in Goodell *et al*. (Goodell *et al*., 1996).

3.3.4 Mammosphere Formation assays

For mammosphere generation, HCC1937 cells were seeded in 96-well low attachment plates (Corning, New York, USA) at the concentration of 1000 cells/well for primary culture, and at 100 cells/well in the following passages. Cells were grown in a serum-free mammary epithelial growth medium (MEBM, Lonza, Verviers, Belgium) supplemented with B27 (Invitrogen, Carsbal, CA), 20 ng/ml EGF, 20 ng/ml bFGF and 4 μg/ml heparin (all from Sigma, St. Louis, MO). Mammosphere were collected by gentle centrifugation (1000 rpm) and dissociated enzimatically (5 min in trypsin/EDTA) and mechanically (26 Gauge needle) and then processed for experiments.

3.3.5 Soft Agar Colony Assay

Anchorage-independent cell growth is one of the hallmarks of cell transformation. The soft agar colony assay is a common method to monitor anchorage-independent growth, which measures proliferation in a semisolid culture media after 3-4 weeks by manual counting of colonies and it is considered the most stringent assay for detecting malignant transformation of cells. Pates were prepared with a coating of 0.75% low-melting agarose (Lonza) by mixing equal volumes of 1.5% agar and 2x growth medium (IMDM). Cells were detached with trypsin, resuspended in growth medium and counted. An overlaid suspension of cells in 0.45% low-melting agarose was prepared by mixing equal volumes of 0.9% agar and 2x growth medium (IMDM) with cells (5×10^4 cells/well). Plates were incubated for 3-4 weeks at 37°C and colonies were then counted under microscope. Experiments were performed in 6-well dishes and also in 12-well dishes to improve the ability of sorted cells to form colonies.

3.3.6 Gene expression analysis by RT-qPCR

Total RNA was prepared using an RNeasy mini kit (Qiagen). RNA was reverse transcribed using the DyNAmo cDNA synthesis kit (FINNZYME) with RNase H+ Reverse Transcriptase and random hexamer primers. Real-time quantitative PCR was carried out using the Universal Probe Library (UPL) technology based on only 165 short hydrolysis probes. Within the human trascriptome, each probe binds to approximately 7000 trascripts but only one specific transcript is detected at a time in a given PCR assay, as defined by the set of chosen PCR primers. In table 6 are reported the UPL probes and the sequences of the primers used for all the analyzed genes. LightCycler TM System instrument was used (Roche Applied Science). Results were displayed as relative levels of genes per *HPRT1*. Each experiment was repeated at three times.

Table 6. UPL probes and primer sequences for analyzed genes

Gene	Primer Forward	Primer Reverse	UPL
Twist 1	ggc-tca-gct-acg-cct-tct-c	cct-tct-ctg-gaa-aca-atg-aca-tct	88
Zeb 1	aac-tgc-tgg-gag-gat-gac-ac	tcc-tgc-ttc-atc-tgc-ctg-a	57
Sox 2	tgc-tgc-ctc-ttt-aag-act-agg-ac	cct-ggg-gct-caa-act-tct-ct	35
Oct 4	ctt-tga-ggc-tct-gca-gct-tag	ctg-ctt-tgc-ata-tct-cct-gaa-g	69
Sox 9	gta-ccc-gca-ctt-gca-caa-c	tcg-ctc-tcg-ttc-aga-agt-ctc	61
Slug	tgg-ttg-ctt-caa-gga-cac-at	gtt-gca-gtg-agg-gca-aga-a	7
FUT 4	cgt-gga-cga-ctt-ccc-aag	gtt-gcg-gtc-gag-gaa-aag	52
E-Cadherin	ccc-ggg-aca-acg-ttt-att-ac	gct-ggc-tca-agt-caa-agt-cc	35
Vimentin	gac-cag-cta-acc-aac-gac-aaa	gga-gca-tct-cct-cct-gca-at	39

3.3.7 Injection in mice and processing of xenografts

All animal experiments were conducted in accord with accepted standards of human animal care. To compare the tumorigenicity of $CD24^{hi}$ *vs* $CD24^{-}$ and of $CD338^{bright}$ *vs* $CD338^{-}$ sorted populations, female NOD/SCID mice aged 4 weeks (Charles River laboratories) were subcutaneously injected. In particular, three NOD/SCID mice were injected with $5x10^5$ $CD24^{hi}$ *vs* $CD24^{-}$ cells into the left and right flanks respectively and two NOD/SCID mice were injected with $5x10^4$ $CD24^{hi}$ *vs* $CD24^{-}$ cells into the left and right flanks respectively. Three NOD/SCID mice were injected with 3 different numbers of $CD338^{+}$ *vs* $CD338^{-}$ cells into the left and right flanks, respectively. To compare the tumorigenicity of $CD338^{bright}$, $CD338^{low}$ and $CD338^{neg}$ sorted populations, *Swiss/nude* mice aged 8 weeks (Charles River laboratories) were injected into a fat pad of mammary gland. In particular, four mice for each of the three populations, were injected with 10^4 cells. In all xenograft experiments, injections of mice with unsorted HCC1937 cells were used as positive controls. Cells from unsorted cell line or sorted cell subpopulations were counted, resuspended in PBS and Matrigel (1:1; BD Biosciences) and injected into mice. Tumor formation was assessed by palpation once a week. Animals were sacrificed as soon as tumor size reached 1.5 cm in diameter or before. Tumor tissues obtained were minced into <1mm pieces, dissociated in an enzymatic solution [CO_2-independent medium complemented with: collagenase A 3 mg/ml (Sigma), penicillin/streptomycin 2%, 100U/ml Hyaluronidase IV, 2% glutamine, 5% FBS and DNase 0.1mg/ml (Roche)] and incubated at 37°C for 60 minutes with gentle agitation. Single cell suspensions obtained were analyzed by flow cytometry after staining with antibodies against the surface antigens of interest.

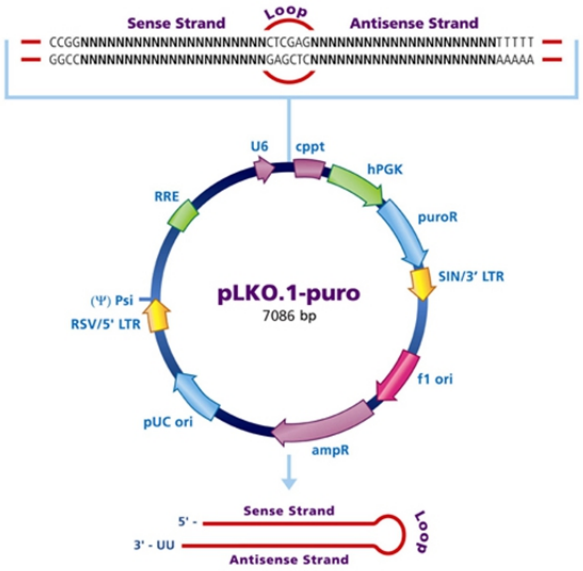
3.3.8 Silencing ABCG2 in HCC1937 cells using lentiviral vector-based shRNA

The strategy of the silencing

To silence ABCG2 I used the MISSION product line of Sigma Alderich that is a viral vector-based RNAi library against mouse and human genes. The libraries consist of bacterial glycerol stocks harboring sequence-verified shRNA lentiviral plasmid vectors for mouse and human genes cloned into the pLKO.1-puro vector (Figure 3.22). MISSION shRNA clones are constructed within the lentivirus plasmid vector, pLKO.1-puro, followed by transformation into *Escherichia coli*. Bacterial cultures, provided as a frozen bacterial glycerol stocks, may be amplified from the glycerol stocks for use in purification of the shRNA plasmid DNA. Subsequently, viral particles can be produced in packaging cells (HEK293T) by co-transfection with compatible packaging plasmids. pLKO.1-puro vectors contain bacterial (ampicillin) and mammalian (puromycin) antibiotic resistance genes for selection of inserts in either bacterial or mammalian cell lines. Upon co-transfection of the plasmids, all required sequences are available to produce and package a viral particle containing the transgene of interest. Only the region between the viral LTRs of the transfer vector is packaged within the viral capsid. Compared to siRNA and other vector-based systems, pLKO.1-puro provides solutions for long-term knockdown and phenotypic observation, transduction of difficult or sensitive cell lines (non-dividing cells or primary cells).

We purchased five MISSION shRNA clones against *ABCG2* from Sigma-Aldrich. These five 21-nucleotide shRNA duplexes from five different parts of the human ABCG2 mRNA (GenBank accession number NM_004827) were designed using the MISSION search database (Zhang *et al*., 2009). The shRNA sequences tested are presented in table 7.

Figure 3.22: Physical map of lentiviral ABCG2 shRNA transfer verctor construct



Name	Description
U6	U6 Promoter
Cppt	Central polypurine tract
Hpgk	Human phosphoglycerate kinase eukaryotic promoter
PuroR	Puromycin resistance gene for mammalian selection
SIN/3' LTR	3' self inactivating long terminal repeat
f1 ori	F1 origin of replication
AmpR	Ampicillin resistance gene for bacterial selection
pUC ori	pUC origin of replication
5' LTR	5' long terminal repeat
Psi	RNA packaging signal
RRE	Rev response element

Table 7: Sequences of shRNA/BCRP clones

A	CCGGGCCTCGATATTCCATCTTCAACTCGAGTTGAAGATGGAATATCGAGGCTTTTTG
B	CCGGGCAACAACTATGACGAATCATCTCGAGATGATTCGTCATAGTTGTTGCTTTTTG
C	CCGGCCTTCTTCGTTATGATGTTTACTCGAGTAAACATCATAACGAAGAAGGTTTTTG
D	CCGGGCTGTGGCATTAAACAGAGAACTCGAGTTCTCTGTTTAATGCCACAGCTTTTTG
E	CCGGCCTGCCAATTTCAAATGTAATCTCGAGATTACATTTGAAATTGGCAGGTTTTTG

Purification of shRNA plasmid DNA

To purify the shRNA plasmids, bacteria from glycerol stocks were cultivated on LB agar to be amplified and to allow to pick single colonies for each shRNA. Single colonies were picked and inoculated into 2ml of LB/ampicillin (100µg/ml) and, 12h after, were transferred into 200ml LB/ampicillin and incubated at 37°C in agitation for over-night. shRNA plasmid DNA was purified using the PureYeld Plasmid Maxiprep System kit (Promega).

Generation of lentiviral particles

Of the many methods to introduce DNA into mammalian cell cultures, the calcium phosphate method is one of the most widely used because it is inexpensive, simple, and suitable for a range of different cell types (Graham & van der Eb, 1973). Lentivirus particles were generated by calcium phosphate co-transfection of 293T cells by three vectors:

1) pLKO.1-puro with one of the five shRNAs
2) pCMVdeltaR8.91 for the expression of the key structural viral packaging genes (Gag-Pol-Tat-Rev)
3) phCMVG-VSVG for the expression of the envelope

The "Mammalian CalPhos transfection kit" of Clontec, providing H_2O sterile, $CaCl_2$ 2M and HBSS 2X, was used to perform the co-trasfection. Particularly, the "transfection mix" was first prepared with: 13.2 µg of pLKO.1-puro with one of the five shRNAs, 10.2 µg of the vector pCMVdR8.91, 2.64 µg of the vector phCMVG-VSVG, 86.8 µl $CaCl_2$, and H_2O to reach a total volume of 700µl. Then, 700 µl of Hank's balanced salt solution (HBSS) were added drop by drop by vortexing the "transfection mix". Transfection mixes were incubated at RT for 25 minutes and then spread on 293T cells. Six hours after transfection, cells were gently washed twice with PBS 1X, and 6ml of fresh medium was added. The day after,

medium was changed with 6ml of fresh one to have at the end, after filtration, 5 ml of virus for infection.

Infection of HCC1937 cells with lentiviral particles and cell selection

Two days after transfection, the viral supernatants were filtred twice with 0.45 µM filters. Polybrene was added (5 µg/ml) to increase the efficiency of infection. It is a cationinc polymer that acts by neutralizing the charge repulsion between the virons and cell surface. The viral supernatants were put on HCC1937 cells (5 ml in 10cm dish). Finally, the medium was changed after six hours. 48h after infection, cells were divided by adding puromycin 1µg/ml to select infected cells, since pLKO.1-puro vector contains the mammalian puromycin antibiotic resistance gene.

3.3.9 Statistical analysis

All measures and experiments were carried out at least three times. Data obtained from mammospheres assays and xenografts experiments were analyzed through t-test and one-way analysis of variance (ANOVA). Differences with values of $p<0.05$ were considered significant. All statistical analyses were carried out using GraphPad Prism software (version 5.00).

IV. PART 2:

Cytometric and biochemical characterization of breast cancer cells from patient-derived breast cancer tissues*

Results from the PART 1 of the present thesis strongly support the presence, in the basal-A BRCA1-mutated cell line HCC1937, of a BCSC subpopulation bearing the phenotype $CD44^{+}CD24^{+}CD338^{bright}$. To conclude that CD338 might constitute a novel antigen reliable to isolate BCSCs from basal-like breast cancers, our results should be extended to human tissues from patients with basal-like breast cancers. Considering this perspective, we carried out a complementary work to standardize the flow-cytomertic methods to study primary cultures established from human breast cancer tissues. While accomplishing this standardization, we have also carried out a wide cytometric and biochemical characterization of cell cultures established from patient-derived breast tissue samples. To evaluate whether cell cultures maintain the cellular heterogeneity found in primary tissues, we compared breast cancer cell cultures from primary tissues with standard breast cancer cell lines through the analysis of expression profiles of a panel of surface markers. We found that *in vitro* culturing of breast cancer cells leads to the selection of myoepithelial/basal breast cancer cells. Furthermore, we demonstrated that breast cancer cell cultures preserve inter-tumor heterogeneity and express stem/progenitor markers that can be identified, quantified and categorized by flow cytometry.

***Leccia F**, Nardone A, Corvigno S, Del Vecchio L, De Placido, Salvatore F, and Veneziani BM. (2012). Cytometric and biochemical characterization of human breast cancer cells reveals heterogeneous myoepithelial phenotypes. Cytometry-A 81(11): 960-972.

4.1 RESULTS

4.1.1 CD44/CD24 expression profile in breast cancer cell cultures

4.1.1.1 Percentages of $CD44^+/CD24^{-/low}$ and $CD44^+/CD24^-$ phenotypes

Cells purified from breast tissues are categorized according to the cell surface expression of CD44 and CD24, which distinguishes $CD44^-/CD24^+$ cells (luminal epithelial cells) from $CD44^+/CD24^-$ cells (basal cells) (Shipstin *et al*., 2007); however, a considerable heterogeneity in CD44 and CD24 expression was reported both between and within breast tumors (Honeth *et al*., 2008). To determine whether these markers could be used to classify luminal and basal breast cancer cells in cell cultures, we analyzed a panel of patient-derived breast cancer cell cultures (n=7) and compared them with standard cell lines previously classified on the basis of gene and protein expression profiling (Neve *et al*., 2006; Shipstin *et al*., 2007; Keller *et al*., 2010). Although all cell cultures displayed a $CD44^+CD24^{-/low}$ phenotype regardless of tumor subtype, biparametric dot plots (Figure 4.1a) of CD44-PE-Cy7 *vs* CD24-AlexaFluor647 show that the prevalence of $CD24^{-/low}$ cells varied among samples. MCF7, MCF10A, cultures #4 and #7 show a low prevalence of this phenotype (Figure 4.1b). Although, from an evaluation of the pattern of the dot plots, it emerges a clear-cut different CD44/CD24 expression pattern between the MCF7 cell line and all the other cell lines and samples, quantitative analysis of $CD44^+CD24^-$ *vs* $CD44^+CD24^+$ cells of cultures #4 and #7 displayed a trend similar to that of MCF7 and MCF10 cells, with a high percentage of $CD24^+$ cells (79.0%, 89.9%, 89.1% and 99.0%, respectively, Table 10 and Figure 4.1b). Thus, the difference of the CD44/CD24 dot plot in MCF7 cell line in comparison to all the other cell lines and samples, is not due to different percentages of $CD44^+/CD24^{-/low}$ cells, but rather to a very low intensity of CD44

expression in MCF7 cells (see section 6.2 of Flow Cytometry Appendix for the difference between percentage and intensity in the analysis of cytometric data).

4.1.1.2 Intensity of CD44 and CD24 expression

We used mean fluorescence intensity (MFI) analyses to estimate the number of CD44 and CD24 molecules per cell (Figure 4.1c,d). CD44 expression was very low in MCF7 cells (MFI = 15, linear scale range: 0-10,000), and high (> 1000) in all the breast cancer cell cultures, as well as in MCF10A, Hs578T and MDAMB231 cells. As shown in Figure 4.1d, intensity of CD24 expression was low in all cell lines and cultures ($1 \leq MFI \leq 109$; MFI linear scale range: 0-10,000). Within this low-range of intensity, CD24 expression was very low in the MCF10A cell line and #5 cell culture (28 and 1 respectively), whereas it was higher in Hs578T and MCF7 cell lines and in culture #4 (MFI> 50). The prevalence of a $CD24^{-/low}$ phenotype and the low intensity of CD24 expression could indicate the *in vitro* loss of luminal phenotype in cell cultures established from patient-derived breast cancer tissues.

4.1.2 Morphology and mammosphere formation

One of the associations in cell line collections is the relationship between the transcriptional phenotype and distinctive biological characteristics such as morphology. Adherent cultures of breast cancer cells had a mixed morphology with few polygonal cells and a prevalence of spindle-like cells (Figure 4.2 upper side of each pair of panels). Mammary gland progenitor cells can be propagated through mammosphere culture. We evaluated whether the patient-derived cell cultures possessed the ability to form spheres. To this aim, we passed the cells (1000/well) to low attachment plates with low serum (0.5% FBS); within 2-3 days the cells formed aggregates that had the characteristic morphology of floating mammospheres (Figure 4.2, lower side of each pair of panels). Fixing the number of seeded cells and the time to 15

days after seeding in non-adherent conditions, if the number of mammospheres formed is similar, we expect that the diameter of the spheres is inversely related to the number of death cells and thus it represents a surrogate index of stem cell survival in non-adherent conditions. We measured the diameter and the number of spheres per well. Although all the cell lines and breast cancer cell cultures formed spheres, the average diameter of aggregates (Figure 4.2, histogram) is heterogeneous, ranging from 50 to 180 µm. Breast cancer cell cultures #4 (mean = 50µm) and #1 (mean = 180µm) formed the smallest and largest size mammospheres, respectively. The mean diameters of the remaining cultures were: culture #2=116 µm, #3=75 µm, #5=100 µm, #6=120 µm, and #7=94 µm.

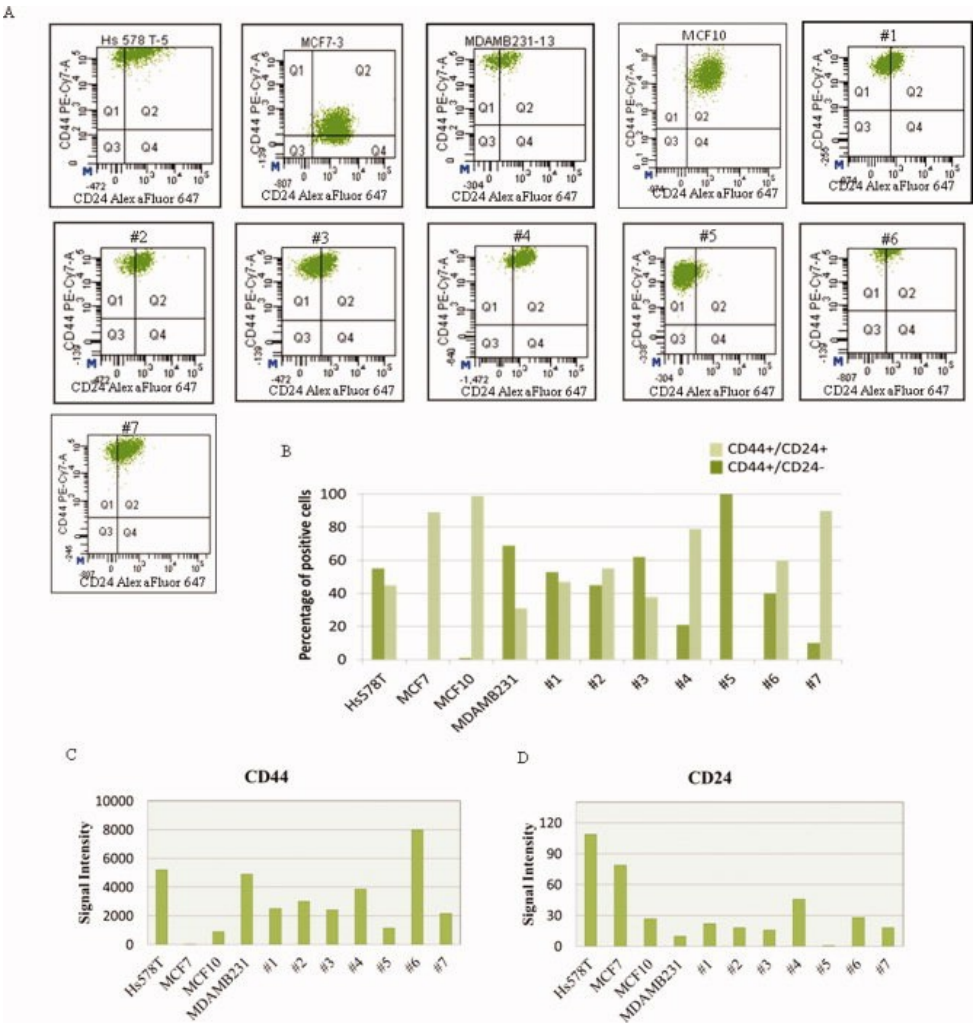


Figure 4.1. CD44 and CD24 expression in breast cancer cell cultures and cell lines. (A) The expression of the two markers is presented on a biparametric dot plot CD44-PE-Cy7 *vs* CD24-AlexaFluor647 for each breast cancer cell culture and cell line. Vertical and horizontal markers delineate the quadrants used to identify the different CD44/CD24 subsets and were set with the appropriate negative control. **(B)** Percentages of $CD44^{+}/CD24^{-}$ (quadrants Q1 in dot plots of panels A) and $CD44^{+}/CD24^{+}$ (quadrants Q2 in dot plots of panels A) cell subpopulations. **(C)** Number of CD44 molecules expressed on the cell surface. The histogram reports the mean fluorescence intensity (MFI, range: 0-10.000) of CD44. **(D)** Number of CD24 molecules expressed on the cell surface of the breast cancer cell cultures and cell lines. The histogram shows the mean fluorescence intensity (MFI, range: 0-10.000) of CD24.

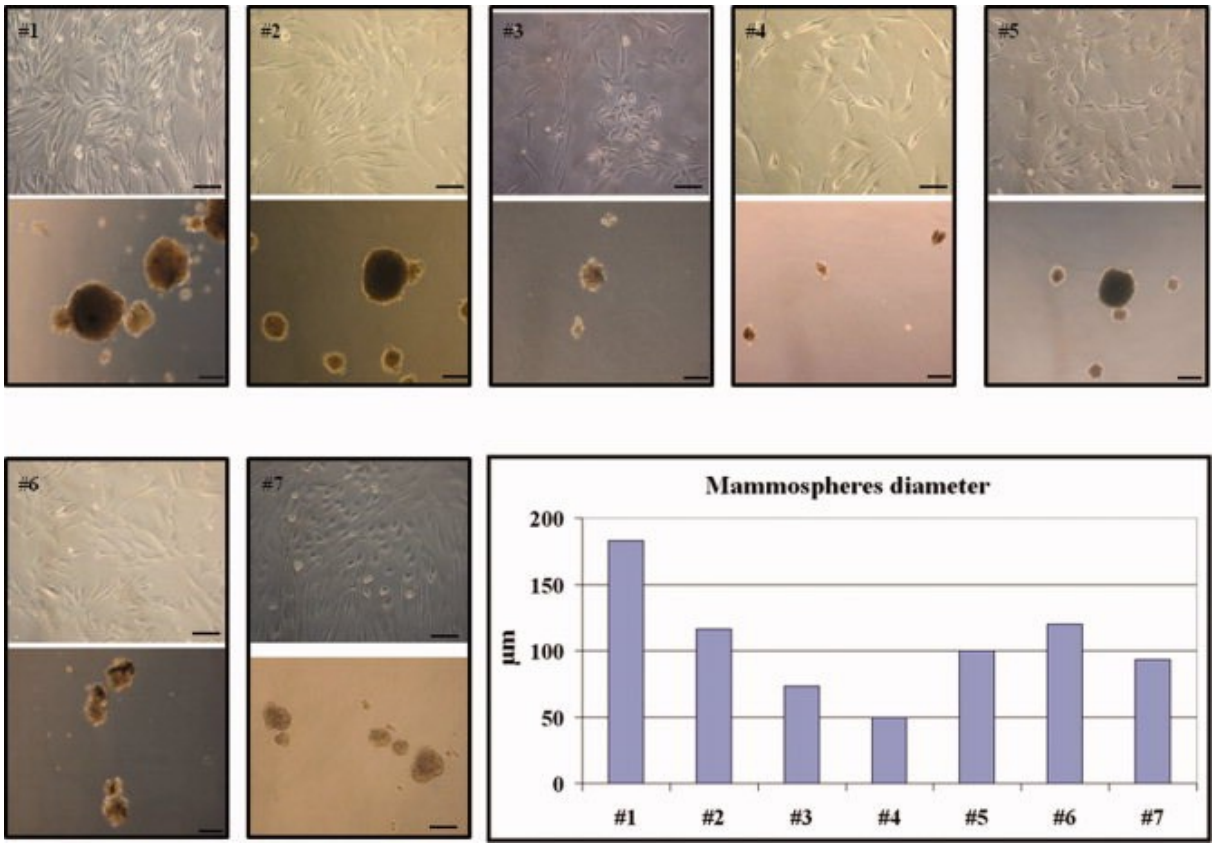


Figure 4.2. Mammospheres formation and size. Phase-contrast microscopy of adherent breast cancer cells #1-7 (upper side of each pair of panels) and the corresponding mammospheres (lower side of each pair of panels). Scale bar 100µm. The graph depicts the average diameter of mammospheres; digital images were analyzed with AxioVision software, the diameter of ten spheres per well was measured and the means are reported in µm. The mean of samples ranged from 50 µm to 150 µm; sample #1, 180 µm; sample #2, 116 µm; sample #3, 75 µm; sample #4, 50 µm; sample #5, 100 µm; sample #6,120 µm, sample #7, 94 µm. The mean standard error was $\leq$ 5%.

4.1.3 Immunophenotypic characterization of breast cancer cell cultures

The 28 surface markers were selected based on their potential to distinguish stem cell populations (CD24, CD44, CD49f, CD133, CD200 and CD326), to mediate cell adhesion and migration (CD9, CD29, CD49b, CD49f, CD61, CD184), metastasis (CD44 and CD184) and signaling (CD66c) (Table 8). We used a four color flow cytometric strategy to measure the expression of these surface markers based on a flow cytometry panel in which cells were stained with MoAbs anti-CD24 and anti-CD44 and with antibodies against all the other analyzed surface antigens (Table 12, Materials and Methods). We used a three gating strategy to define the target cell population which should be analyzed for the expression of surface markers. (Figure 4.3).

Tables 9 and 10 show the percentages of cells expressing surface markers in all breast cancer cell lines and patient-derived cultures. All markers showed a unimodal profile in all cell lines and cell cultures, which is indicative of homogeneity of a cellular population within the sample (see section 6.2 of Flow Cytometry Appenix). An exception to the general unimodal pattern was the expression of CD105 (endoglin) in sample #1 (Figure 4.3). Ten out of 28 surface markers were homogeneously expressed, at high or low level, in all cell lines and cell cultures (Table 9). These were CD29 (β1 integrin), CD55 (DAF), CD59 (MIRL), CD66b (CEACAM8), CD66c (CEACAM6), CD81 (TAPA1), CD151 (PETA-3), CD165 (AD2), CD166 (ALCAM) and CD324 (E-cadherin). On the other hand, 18/28 surface markers were heterogeneously expressed (Table 10). CD90 (Thy-1), which is expressed on normal basal cells but not on luminal cells (Donnenberg *et al*., 2010), was highly expressed in all cell cultures and in the Hs578T cell line (100% of positive cells), not expressed in the luminal MCF7 cell line, and intermediately expressed (60.4%) in the basal MCF10A cells. CD10 (CALLA), a marker of myoepithelial cells, was highly expressed in all cell cultures (100% of positive cells) and poorly expressed in the MCF7 and MDAMB231 cell lines. CD326

(EpCAM) and CD227 (MUC-1), two key markers of luminal epithelial cells, were highly expressed in the luminal MCF7 cell line, 100% and 85.3% of positive cells respectively. Conversely, CD326 and CD227 expression was low or absent in the Hs578T and MDAMB231 basal-B/mesemchymal cell lines and intermediate in the MCF10A cell line. In all patient-derived cell cultures, CD326 was expressed at very low levels, whereas the expression of CD227 was 42.4% and 40.0% for samples #2 and #7, respectively. CD54 (ICAM-1), a molecule whose function is required for invasion of metastatic breast cancer cells, is highly expressed either in the cell cultures (from 85.1% to 99.7% of positive cells) and in the cell lines (MCF7=94.7%, MDA=83%, HS578T=100%, MCF10A=64.0%), although differences are greater than 5%, and we thus considered it as a heterogeneously expressed marker. In conclusion, immunophenotipic characterization of patient-derived breast cancer cell culture confirm the *in vitro* loss of luminal phenotype, as also suggested by the CD44/CD24 expression profile (section 4.1.1).

Table 8. Molecular identity and function of the cell surface markers analyzed in breast cancer cell lines and cell culturer from patient-derived breast tissue samples.

CD	Molecule	Function	References
CD9	P24	Cell adhesion and migration	Nishida 2009
CD10	CALLA	Antigen overexpressed on many tumors	Iwaya 2002
CD24	HSA	Adhesion and spread of metastatic tumors	Al-Hajj 2003
CD26	DPPIV	Exopeptidase, tissue restructuring	Pro 2004
CD29	β1 integrin	Adhesion to matrix proteins	Charafe-Jauffret 2009
CD44	H-CAM	Sppression of apoptosis, metastasis	Al-Hajj 2003
CD47	Rh-associated protein	Cell activation, apoptosis, cell spreading	Manna 2004
CD49b	α2 integrin	Cell adhesion to collagen and laminin	Langsenlehner 2006
CD49f	α6 integrin	Cell surface signaling	Cariati 2008
CD54	ICAM	Cell adhesion, immune reactions	Rosette 2005
CD55	DAF	Protection against complement	Ikeda 2008
CD59	MIRL	Protection from complement-mediated lysis	Babiker 2005
CD61	β3 integrin	Cell-surface mediated signaling	Charafe-Jauffret 2009
CD66b	CEACAM8	Cell adhesion, transmembrane signalling	Lasa 2008
CD66c	CEACAM6	Cell adhesion, transmembrane signalling	Lasa 2008
CD81	TAPA-1	Response to antigens	Yáñez-Mó 2009
CD90	Thy-1	Proliferation and differentiation	Yamazaki 2009
CD105	ENG (Endoglin)	Angiogenesis, vessel wall integrity	Henry 2011
CD133	Prominin 1	Unknown	Wright 2008
CD151	PETA-3	Cell adhesion	Sadej 2009
CD164	MGC-24	Adhesion and homing	Havens 2006
CD165	AD2	Unknown	Seon 1984
CD166	ALCAM	Adhesion, organ development	Kulasingam 2009
CD184	CXCR4	Increased expression in mammospheres	Krohn 2009
CD200	OX2	Immunosuppression	Kawasaki 2007
CD227	MUC1	Response to hormones and cytokines	Stingl 2009
CD324	E-Cadherin	Cell adhesion, tumor suppression	Prasad 2009
CD326	EpCAM	Cell adhesion	Stingl 2009

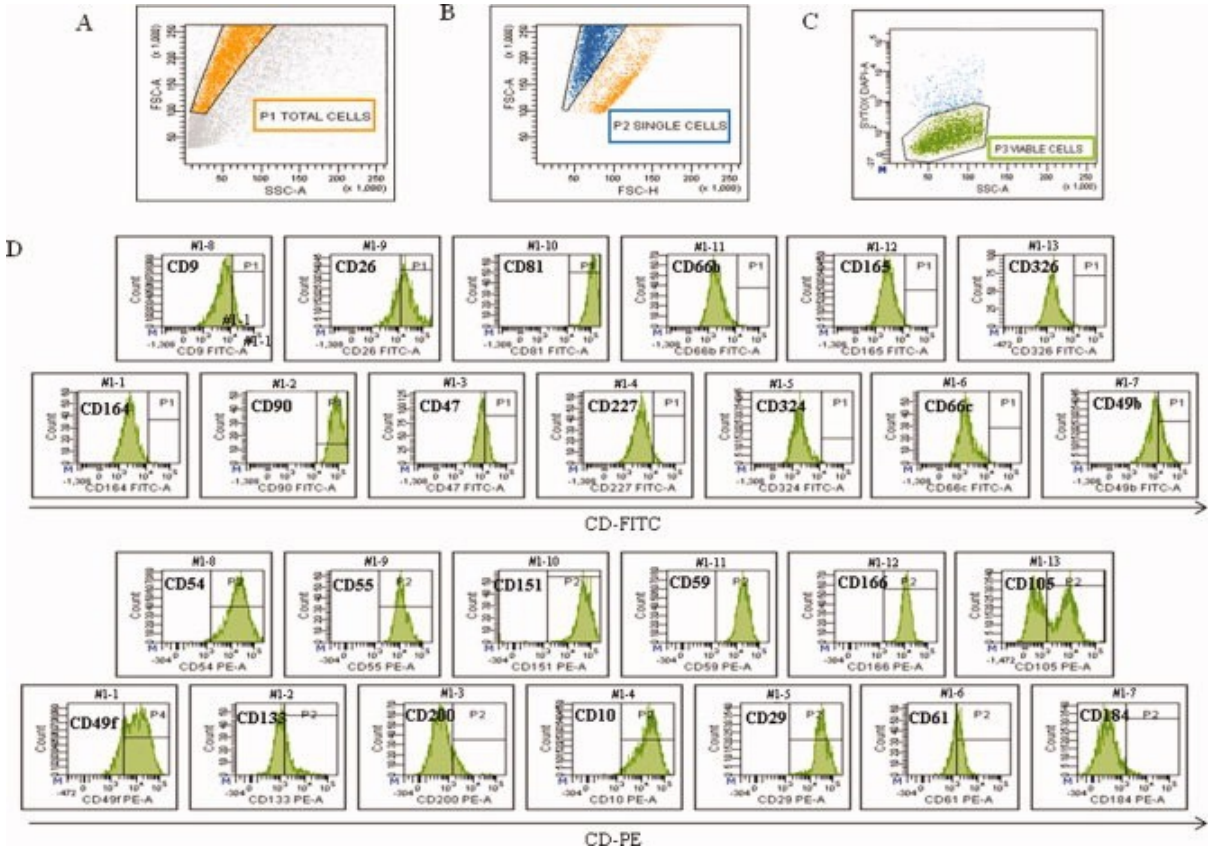


Figure 4.3. Representative strategy of flow cytometry analysis of breast cancer cell culture #1. **(A,B,C)**: gating strategy used to remove debris, cell clusters and dead cells. **(A)** To exclude dead cells and debris, cells were gated on a two physical parameters dot plot measuring forward scatter (FSC) *vs* side scatter (SSC) (see section 6.2 of Flow Cytometry Appendix); **(B)** Then doublets were excluded by gating cells on FSC-Height *vs* FSC-Area dot plots (see section 6.2 of Flow Cytometry Appendix)**; (C)** Finally, Sytox Blue negative cells were gated. **(D)** Surface marker expression on cells gated as described of #1 breast cancer cell culture. The expression of each antigen is represented on a frequency distribution histogram (count *vs* FITC or PE signal) (see section 6.2 of Flow Cytometry Appendix). The vertical marker on each histogram used to detect the antibody-positive cells was established using the appropriate negative controls.

Table 9. Antigens homogeneously expressed* in breast cancer cell lines and cell cultures from patient-derived breast tissue samples

Negative or very lowly expressed antigens†

	Cell line				**Breast cancer cell culture**						
CD	**MCF7**	**MDAMB231**	**Hs578T**	**MCF10**	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**
66b	0.1‡	1.4	0.7	1.7	0.4	0.0	0.3	0.3	0.0	0.0	0.8
66c	0.4	1.5	0.7	4.0	0.9	0.0	0.2	3.5	0.0	0.0	1.2
165	0.1	0.4	3.9	1.9	1.6	0.4	0.6	1.5	4.0	7.0	6.5
324	4.2	2.1	3.7	2.5	0.5	3.3	0.5	1.2	0.4	0.0	7.0

Highly expressed antigens†

	Cell line				**Breast cancer cell culture**						
CD	**MCF7**	**MDAMB231**	**Hs578T**	**MCF10**	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**
29	100.0	100.0	99.9	100.0	100.0	100.0	99.9	100.0	99.7	100.0	100.0
55	100.0	99.1	99.8	99.5	100.0	98.8	99.2	99.6	96.2	100.0	100.0
59	100.0	99.8	98.9	100.0	100.0	100.0	100.0	100.0	99.6	100.0	100.0
81	100.0	99.4	99.5	98.7	100.0	100.0	100.0	99.6	99.8	100.0	99.9
151	100.0	99.8	99.8	94.9	99.5	99.8	100.0	100.0	99.8	100.0	100.0
166	100.0	100.0	100.0	90.1	100.0	100.0	98.9	100.0	100.0	100.0	98.5

*Molecules with expression differences $\leq 5\%$ were considered homogeneously expressed, whereas those with differences $> 5\%$ were considered heterogeneously expressed.

†Negative or expressed at very low level: $\leq 7\%$; highly expressed : $\geq 99.1\%$

‡Percentage of cells within each cell line or primary culture that were antibody-positive and therefore expressing the indicated antigens. The percentages of antibody-positive cells were calculated by comparison with the appropriate negative control.

Table 10. Antigens heterogeneously expressed* in breast cancer cell lines and cell cultures from patient-derived breast tissue samples

CD	MCF7	MDA MB231	Hs578T	MCF10	#1	#2	#3	#4	#5	#6	#7
9	99.9 †	55.1	0.2	99.5	16.6	74.4	9.5	2.8	42.2	36.7	98.0
10	10.7	16.6	99.1	93.1	99.2	100.0	100.0	99.6	100.0	98.0	100.0
24	89.1	31.1	45.5	99.0	47.5	54.6	37.3	79.0	0.0	60.0	89.9
26	0.2	1.1	79.3	4.4	69.8	74.2	56.0	8.3	2.6	50.9	88.8
44	71.0	98.0	100.0	99.9	100.0	99.8	100.0	100.0	100.0	100.0	100.0
47	100.0	71.6	60.0	99.5	21.7	77.4	60.0	88.2	82.0	69.0	99.9
49b	99.9	100.0	42.0	99.9	39.9	13.5	93.7	96.9	79.4	94.0	99.8
49f	21.2	99.5	91.1	100.0	82.1	74.7	15.2	89.8	4.3	12.2	92.0
54	94.7	83.0	100.0	64.0	99.4	85.1	92.7	99.7	85.4	91.0	97.7
61	0.2	59.9	41.5	27.0	58.4	23.0	93.5	74.2	89.1	50.0	65.7
90	0.3	6.7	100.0	60.4	100.0	100.0	100.0	99.5	100.0	100.0	100.0
105	22.2	61.2	98.4	38.4	67.5	100.0	21.9	99.7	97.0	83.0	23.9
133	72.0	22.6	92.9	18.0	33.3	45.9	13.9	1.8	41.3	90.0	1.5
164	33.0	69.7	0.4	24.2	0.6	5.1	0.8	0.4	4.3	0.0	4.0
184	0.9	11.6	12.6	41.0	2.4	11.9	5.0	1.7	7.3	6.0	11.0
200	4.4	24.5	26.4	90.9	12.0	0.7	0.2	6.7	6.7	20.0	24.0
227	85.3	1.7	2.3	33.8	3.2	42.4	3.3	3.2	0.0	2.0	40.0
326	100.0	1.8	4.0	77.8	0.6	0.1	0.7	0.7	0.0	0.0	0.4

*Molecules with expression differences $\leq 5\%$ were considered homogeneously expressed, whereas those with differences $> 5\%$ were considered heterogeneously expressed.

†Percentage of cells within each cell line or cell culture that were antibody-positive and therefore expressing the indicated antigens. The percentages of antibody-positive cells were calculated by comparison with the appropriate negative control.

4.1.4 Cell culture phenotype correlates with the myoepithelial/mesenchymal phenotype

We used the non-parametric Spearman correlation test to compare the phenotypic heterogeneity of breast cancer cell cultures with that of control cell lines (Table 11). Although displaying heterogeneous expression of surface markers, cell cultures were strongly correlated with Spearman correlation coefficients (rs), which ranged from 0.7 to 0.9 ($0.7 < rs > 0.9$). All cell cultures were also strongly correlated with the cell line Hs578T ($0.6 < rs > 0.9$) and, even if to a lesser extent, to the MDAMB231 cell line and to the MCF10A ($0.1 < rs > 0.6$) cell line. They were not correlated with the MCF7 cell line ($-0.3 < rs > -0.04$, $p>0.05$). Since Hs578T and MDAMB231 are basal-B/mesenchymal cell lines, whereas MCF7 is a luminal cell line, these results further support the selection of basal/mesenchymal phenotype in patient-derived breast cancer cell cultures.

4.1.5 Expression of epithelial/mesenchymal markers

To determine whether patient-derived cultured cells expressed the epithelial or mesenchymal phenotype we immunoblotted for cytokeratins (CK), vimentin and SMA (Figure 4.4). CK18 and CK19 are epithelial markers, whereas CK5, SMA and vimentin are myoepithelial/mesenchymal markers (Blick *et al*., 2008). Densitometric analysis, reported as percentage increase *vs* tubulin, showed that CK18 and CK19 were highly expressed, more than 0.5% increase, in MCF7 cells (2.08 and 1.9) and in culture #4 (1.05 and 1.3) (Figure 4.4). CK5, SMA and vimentin were not expressed in MCF7 epithelial cells but present in the Hs578T, MDAMB231, MCF10 cell lines, and all the breast cancer cultures. Cultures #1, #4 and #6 expressed high levels of SMA (>0.5% increase).

Table 11. Correlations between human breast cancer cell lines and cell cultures based on heterogeneously expressed surface markers

	MCF7	**MDA MB231**	**Hs578T**	**MCF10**	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**
MDA MB231	0.316*									
Hs578T	-0.215	0.214								
MCF10	0.443	0.577	0.125							
#1	-0.321	0.257	0.893	0.223						
#2	-0.096	0.187	0.778	0.225	0.817					
#3	-0.158	0.299	0.735	0.260	0.839	0.699				
#4	-0.065	0.463	0.822	0.429	0.872	0.799	0.798			
#5	-0.204	0.349	0.639	0.178	0.643	0.675	0.744	0.697		
#6	-0.042	0.270	0.831	0.244	0.801	0.724	0.867	0.802	0.779	
#7	0.034	0.334	0.512	0.604	0.699	0.689	0.807	0.728	0.604	0.707

*Nonparametric Spearman correlation coefficient (rs).

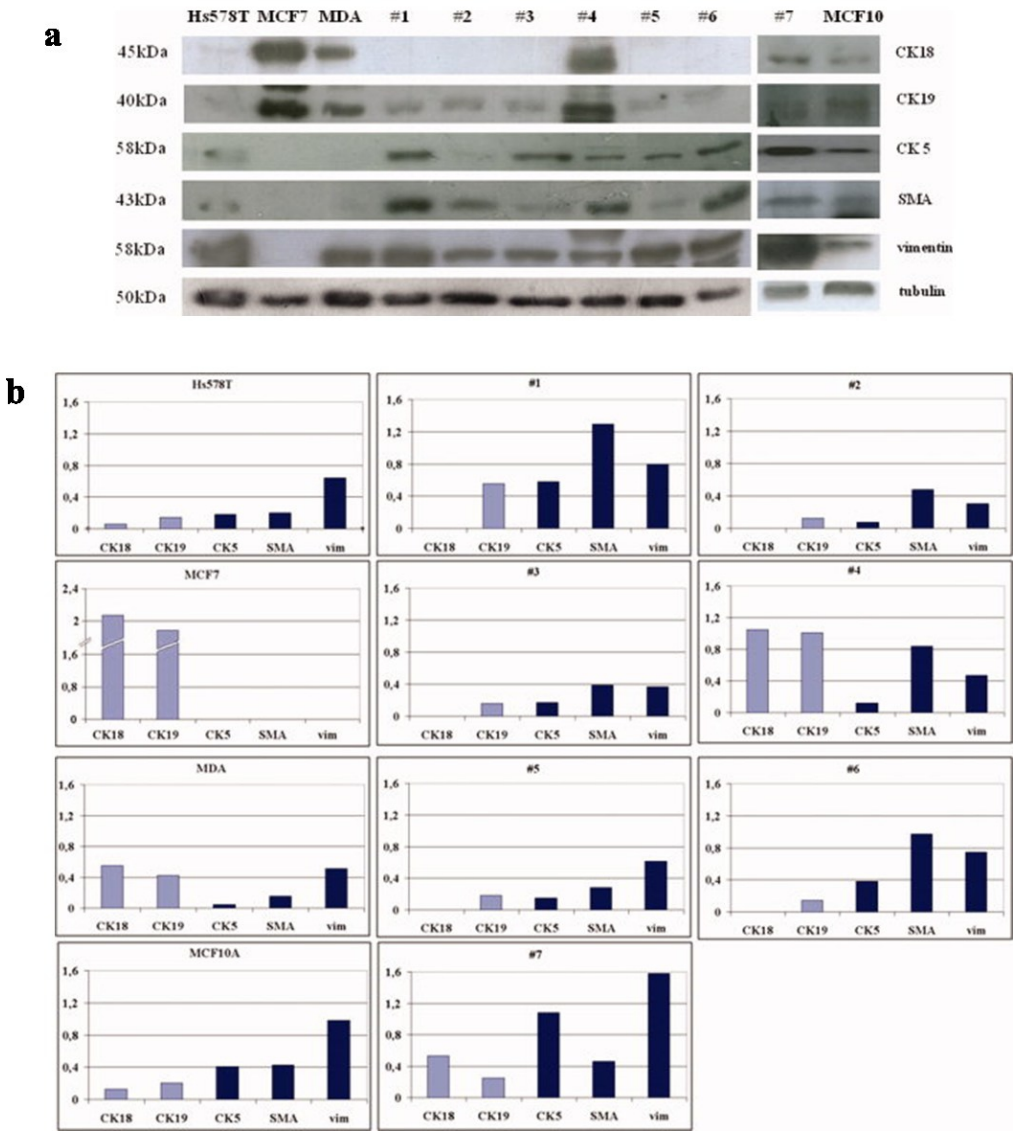


Figure 4.4. Analysis of luminal and basal markers expression in breast cancer cell lines and cell cultures from patient-derived breast tissue samples. (a) Western blot of cytokeratin 18 (CK18), cytokeratin 19 (CK19), cytokeratin 5 (CK5), alpha smooth muscle actin (SMA) and vimentin (vim) in control cells Hs578T, MCF7, MDA-MB231, MCF10A and in sample cells #1, #2, #3, #4, #5, #6, #7. 30 µg of total protein extract were loaded. Tubulin was used as a loading control. **(b)** Densitometric analysis with Scion Image was performed with cytokeratin 18 (CK18) and cytokeratin 19 (CK19) (light bar), cytokeratin 5 (CK5), alpha smooth muscle actin (SMA), vimentin (Vim) (dark bar), in control cells Hs578T, MCF7, MDA-MB231, MCF10A, and sample cells #1, #2, #3, #4, #5, #6, #7. Data were normalized for tubulin; standard error, not reported on graph, was $\leq$ 10%.

4.2 DISCUSSION

In vitro culture of human breast cancer cells usually leads to selection for basal/myoepithelial (Neve *et al*., 2006; Keller *et al*., 2010). We have immunophenotyped breast cancer cell cultures and cell lines for 28 surface markers to analyze the cell surface phenotype, and showed that a panel of CD antigens can be used to determine inter-tumor heterogeneity.

To determine the cellular heterogeneity of tumors, we used a four-color flow cytometric strategy in which CD24 and CD44 stained cells were also stained with MoAb against two other antigens (Table 12, Materials and Methods section). The 28 surface markers were selected based on the potential to distinguish between cell types of epithelial or mesenchymal lineage or to identify putative cancer stem/progenitor cells. Every marker showed a unimodal profile in all samples, which indicates the lack of cellular subpopulations. The only exception to this general pattern is the expression of CD105 (endoglin) in sample #1, the significance of which requires further investigation. Flow cytometric analysis showed that 10 surface markers were homogeneously expressed in all cell lines and breast cancer cell cultures. Between these ten surface markers, some, such as CD324 (E-cadherin) displayed very low values, and others, such as CD166 (ALCAM), displayed high values. Particularly, four markers (CD66b, CD66c, CD165, CD324) displayed negative/low expression in all the cell lines and breast cancer cells tested. Every marker is involved in cell adhesion. CD66b and CD66c MoAbs react with CEACAM-8 and CEACAM-6, respectively. Members of the CEACAM family play a role in invasion and metastasis, are involved with cellular differentiation (Ilantzis *et al*., 2002), and are expressed in breast cancer carcinomas (Esteban *et al*., 1994). CD324/E-cadherin. E-cadherin functions as a tumor suppressor gene. In breast cancer, loss of E-cadherin expression is a necessary component for invasion and metastasis (Peinado *et al*.,

2007), and it is a marker of epithelial-mesenchymal transition (Blick *et al*., 2008). Reduction of CD324 (E-cadherin) expression is associated with breast carcinomas of basal-like and triple negative phenotype (Mahler-Araujo *et al*., 2008) and with the acquisition of cancer stem cell associated characteristics like increased $CD44^{+}/CD24^{-}$ ratio (Sigurdsson *et al*., 2011). On the othe hand, among the ten homogeneously expressed markers, six displayed homogenous high expression in all the cells, namely CD29 (beta1-integrin), CD55 (DAF), CD59 (MIRL), CD81 (TAPA1), CD151 (PETA-3) and CD166 (ALCAM). CD55 (DAF) and CD59 (MIRL) are two complement regulatory proteins expressed on cells to avoid autologous complement attack. In breast cancer, there is a high variability in complement regulatory protein expression, with some tumors expressing only one inhibitor, and others expressing various combinations of two or three inhibitors (Niehans *et al*., 1996). CD81 (TAPA1) and CD151 (PETA-3) belong to the four-transmembrane domain (tetraspanin) family of proteins involved in regulating tumor cell motility and invasiveness, mainly through their effect on the adhesive and signaling function of integrin (Yànez-Mo *et al*., 2009). Integrin-tetraspanin complexes modulate tumor cell-cell and cell-substrate interactions (Stipp *et al*., 2010; Romanaska *et al*., 2011). CD166 (ALCAM) is overexpressed in breast cancer metastases (Ihnen *et al*., 2011). High values of CD166 are associated with high values of CD44 and either high or low values of CD24 (Stuelten *et al*., 2010).

The homogeneous, low or high, expression of surface markers suggests that cell lines and breast cancer cell cultures share many of the molecular features of their acquired ability to grow *in vitro*, and acquired adaptation to culture conditions. This is in agreement with previous reports (Neve *et al*., 2006; Keller *et al*., 2010; Petersen *et al*., 2010). While the markers in the homogenously expressed group (Table 9, Results section) could be considered molecules ubiquitously expressed for the biological adaptation of cells to *in vitro* conditions, the heterogeneously expressed group (Table 10, Results section) is likely to consist of

molecules that identify cells of different origin. In agreement with others (Keller *et al*., 2010), we found that heterogeneity within each cell line and patient-derived breast cancer cell culture was remarkably restricted (unimodal profile), which could indicate that *in vitro* culturing enriches for cellular phenotypes that may represent the progeny of different stem/progenitor cells. On the other hand, we found that 18/28 antigens were heterogeneously expressed between all breast cancer cell lines and cultures. If the different distribution of marker expression is influenced by the culture conditions, then the expression profile patterns in the cell lines and breast cancer cell cultures would appear not to be related. We thus performed correlations using Spearman's rank, on the base of heterogeneously expressed markers. We found no correlation of the patient-derived breast cancer cells with MCF7 cell line, a weak to moderate correlation with MDA-MB231 and with MCF10A, statistically significant correlations ($p \leq 0.001$) with Hs578T and a strong statistically significant correlation within the different breast cancer cell cultures (Table 11, Results section). Since heterogeneity between breast cancer cell cultures does not depend on culture conditions because all the primary cultures undergo the same protocol of isolation, the flow cytometry analysis shows the potential to identify individual features of a breast tumor.

Studies designed to define the cell of origin in experimental model systems report that mesenchymal stem cells, isolated from various normal and pathological tissues, show phenotypic heterogeneity (Chamberlain *et al*., 2007) and it has been suggested that they reside in virtually all postnatal tissues. The resulting cultures are morphologically heterogeneous; in fact, they contain cells ranging from spindle-shaped cells to polygonal cells and some cuboidal cells. Mesenchymal stem cells express a number of markers, none of which are specific to mesenchymal stem cells. Although many of these markers display variable expression due to differences in (i) tissue source, (ii) the method of isolation and culture, and (iii) species, it is generally agreed that adult human mesenchymal stem cells can express

CD44 (H-CAM), CD90 (Thy-1), CD105 (Endoglin), CD184 (CXCR4), various integrin molecules, such as CD49b (α2integrin), CD49f (α6integrin), CD29 (β1integrin), CD61 (β3integrin), and other adhesion molecules including CD54 (ICAM) and CD166 (ALCAM) (Chamberlain *et al*., 2007). Within our breast cancer cell cultures, which were all $CD44^{+}$, concomitant high expression of surface markers associated with the mesenchymal stem cell phenotype included CD10/CALLA, CD54/ICAM and CD90/Thy-1, and was paralleled by low expression of CD326-EpCAM.

Breast cancers are broadly classified histopathologically by the expression of either luminal cytokeratins (CK18/19) or basal cytokeratins (CK5), SMA and vimentin (Sorlie *et al*., 2001); the categorization in subtypes is critical for the prediction of clinical outcome and treatment response. To refine the portrait of patient-derived breast cancer cell cultures, we examined whether the markers associated with myoepithelial/basal-like breast phenotype (CK5, SMA and vimentin) and those associated with luminal phenotype (CK18 and CK19) were differentially expressed based on culture conditions. Particularly, myoepithelial cells are cells from the basal layer, defined as cells expressing both epithelial markers and contractile proteins (Anbazhagan *et al*., 1998). These cells are distinguished from basal cells in multilayered squamous epithelium because they express proteins characteristics of mesenchymal cells, such as vimentin (Rønnov-Jessen *et al*., 1995), alpha SMA (Gusterson *et al*., 19982), and cytokeratins 5 (CK5, or basal cytokeratin) (Nagle *et al*., 1986). However, studies aimed at defining the cell of origin in breast tumors show that the expression of CK5 is not restricted to myoepithelial cells in tissue cultures (Gusterson *et al*., 2005). We used western blot to evaluate the expression of these markers in all cell lines and breast cancer cell cultures. Although the overall levels of marker expression were different, all of the breast cancer cell cultures analyzed in this study expressed CK5, SMA and vimentin, all markers of basal/myoepithelial cells, as well as the basal-b/mesenchymal cell lines MDAMB231 and

Hs578T, whereas MCF7 do not. The majority of human breast cancers cell lines, were cytokeratin 18 (CK18) and cytokeratin 19 (CK19) positive, although expressing them at different levels. Conversely, all the breast cell cultures, excluding culture #4, were negative or weakly positive for CK19 and CK18. Taken together, results of immunophenotype and western blot analysis indicate that the patient-derived breast cancer cell cultures displayed a myoepithelial/mesenchymal phenotype. Our results, are in agreement with the *in vitro* loss of phenotype restricted to the luminal lineage (Keller *et al*., 2010).

4.3 MATERIALS AND METHODS

4.3.1 Materials

Standard medium consisted of minimal essential Dulbecco/Ham F12 (1:1) (DMEM/F12) (Sigma-Aldrich), supplemented with 2 mM glutamine (Sigma-Aldrich), P/S, 15 mM HEPES (Sigma-Aldrich) and 5% FBS. MCF-7, MDA-MB231, Hs578T and MCF10A were from ATCC (American Type Culture Collection). Multi-color flow cytometry was performed with anti-human monoclonal antibodies (MoAbs) that were conjugated with phycoerythrin (PE), fluorescein isothiocyanate (FITC), phycoerythin-Cy7 (PE-Cy7) or Alexa Fluor 647. PE-conjugated MoAbs against CD10, CD29, CD49f , CD54, CD55, CD59, CD61, CD66c, CD151, CD166 and CD200, and FITC-conjugated MoAbs against CD9, CD26, CD47,CD49b, CD66b, CD81, CD90, CD164, CD165, CD227, CD324 and CD326 were from BD Biosciences and BD Pharmingen (San Jose, CA, USA); PE-conjugated MoAb against CD133 from Miltenyi Biotech (Auburn, CA, USA); AlexaFluor647-conjugated MoAb against CD24, and PE-Cy7-conjugated MoAb against CD44 were from Biolegend (San Diego, CA, USA); PE-conjugated MoAb against CD184 was from Immunotech (Marseille, France); PE-conjugated MoAb anti-CD105 was from Serotec (Kidlington, Oxford, UK). For western blot analysis the following MoAb CK5 (AE14) cod: sc-80606, CK19 (BA17) cod: sc-53258 and SMA (alpha Smooth Muscle Actin) (CGA7) cod: sc-53015 were diluted 1:200. The secondary antimouse antisera (sc-2005) were diluted 1:3000. The polyclonal antibodies used were as follows: pNeu (Tyr-1248)-R cod: sc-12352, diluted 1:100, CK18 (H-80) cod: sc-28264 diluted 1:200, ER-alpha (HC-20) cod: sc-543 diluted 1:400 and EGFR (1005) cod: sc-03 diluted 1:1000. The secondary anti-rabbit antisera (sc-2004) were diluted 1:3000. All the antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). The monoclonal anti-Tubulin (clone DM 1A) code: T-9026 diluted 1:400 and vimentin (clone V9)

code: V-6630 diluted 1:1000 were from Sigma-Aldrich (Milan, Italy). All dilutions were made in TBS-Tween 20 (Thermo Fisher Scientific, Milan, Italy) with 5% non fat milk (Applichem, Darmstadt, Germany).

4.3.2 Breast tumor specimens and cell culture conditions

Tumor histotype, size, grading and markers including ER were determined with standard procedures, and HER2 was determined with HercepTest (Dako, Carpintera, CA, USA). The following data were retrieved from the pathology reports of each patient: patient #1, invasive lobular carcinoma (ILC), pT_1 N_0 (T = tumor size, N = lymph node involvement), ER-positive (Estrogen Receptor alpha) (80%) and without c-erbB2 over-expression; patient #2, invasive ductal carcinoma (IDC) with a mucinous component, pT_2 N_0, ERα-positive (80%) and c-erbB2-negative; patient #3, IDC, pT_1 N_1, ERα-negative (<5%) and c-erbB2 overexpression (+++); patient #4, IDC, pT_1 N_1, ERα-positive (80%), and c-erbB2 overexpression (+++). Patient #5, IDC, pT_2 N_0, and ERα-negative, without c-erbB2 over-expression; patient #6, IDC arising from a contralateral recurrence, pT_{1c} N_x, ERα-positive (>90%) and c-erbB2-negative; patient #7, IDC, pT_1, N_0, ERα-negative (<5%) and c-erbB2-negative. All patients had a high grade of malignancy.

Breast cancer tissue specimens were collected, processed, cultured and frozen as cell suspension via standardized operative procedures for banking (Veneziani *et al*., 2007) and cell population expansion (Nardone *et al*., 2011). Control MCF-7, MDA-MB231, Hs578T and MCF10A were cultured in DMEM plus 10% FBS. Frozen cells were thawed, allowed to adhere and harvested 15-20 days in standard medium.

4.3.3 Mammosphere formation efficiency and size

Cells were dissociated and seeded, 1000 cells/well, in ultra-low attachment surface 24 well plates (Corning) in DMEM/F12 plus 0.5% FBS medium, according to Dontu *et al*. The medium was renewed twice weekly. Mammospheres were cultured for 15 days and measured under an inverted microscope Axiovert 40 C (Zeiss, Milan, Italy) equipped with a Canon powershot A640 camera (Zeiss). Digital images were analyzed with AxioVision software (Zeiss).

4.3.4 Flow cytometry analysis

Cell suspensions were analyzed for the expression of 28 surface antigens, namely, CD9 (p24), CD10 (CALLA), CD24 (HSA), CD26 (DPPIV), CD29 (β1 integrin), CD44 (H-CAM), CD47 (Rh-associated protein), CD49b (α2 integrin), CD49f (α6 integrin), CD54 (ICAM), CD55 (DAF), CD59 (MIRL), CD61 (β3 integrin), CD66b (CEACAM8), CD66c (CEACAM6), CD81 (TAPA1), CD90 (Thy1), CD105 (Endoglin), CD133 (prominin1), CD151 (PETA-3), CD164 (MUC-24), CD165 (AD2), CD166 (ALCAM), CD184 (CXCR4), CD200 (OX2), CD227 (MUC1), CD324 (E-Cadherin), CD326 (EpCAM) (Table 8, Results section). We used a four color flow cytometric strategy to measure the expression of these surface markers (Table 12). Briefly, each experiment consisted of 13 tubes, each containing MoAb anti-CD24-AlexaFluor647, MoAb anti-CD44-PE-Cy7, one PE-conjugated and one FITC-conjugated antibody against two of the other CD antigens to be analyzed. For all experiments, after enzymatic detachment from non-saturated cultures, cells were counted, resuspended at $5x10^5$ in 100µl of PBS and stained by incubation at 4°C for 20 min with the appropriate amount of above the MoAbs in PBS. After staining, all samples were washed twice with PBS, pelletted and suspended in 0.5 ml of FACS buffer (FACS Flow Sheat Fluid, BD Biosciences) for FACS analysis. A few minutes before FACS acquisition, cells were incubated at room

temperature in the dark with a vital dye (SytoxBlue, Invitrogen) to exclude dead cells from the analysis. In all experiments, for each cell type, a negative control was treated with the same procedure but without antibody staining. This negative control represents the background caused by the cellular autofluorescence (see section 6.2 of Flow Cytometry Appendix). The samples were analyzed with a FACSAria flow cytometer and the FACSDiva software (Becton Dickinson, Franklin Lakes, NJ, USA). For each sample run, 10,000 to 20,000 events were recorded and analyzed. We used a three gating strategy that is the same described in figure 4.3.

Table 12. Four-color flow cytometry panel for expression analysis of surface markers in breast cancer cell lines and cell cultures from patient-derived breast tissue samples

Tube	FITC	PE	PE/Cy7	AlexaFluor647
1	CD164	CD49f	CD44	CD24
2	CD90	CD133	CD44	CD24
3	CD47	CD200	CD44	CD24
4	CD227	CD10	CD44	CD24
5	CD324	CD29	CD44	CD24
6	CD66c	CD61	CD44	CD24
7	CD49b	CD184	CD44	CD24
8	CD9	CD54	CD44	CD24
9	CD26	CD55	CD44	CD24
10	CD81	CD151	CD44	CD24
11	CD66b	CD59	CD44	CD24
12	CD165	CD166	CD44	CD24
13	CD326	CD105	CD44	CD24

Each experiment consisted of 13 tubes. Each tube contained the MoAb anti-CD24-AlexaFluor647, the MoAb anti-CD44-PE-Cy7, one PE-conjugated and one FITC-conjugated antibody against two of all the other CD antigens to be analyzed.

4.3.5 Western blot analysis

Whole cell extracts were obtained by lysing samples in 50 mmol/L of TRIS (pH 7.5), 100 mmol/L of NaCl, 1% NP40, 0.1% Triton, 2 mmol/L of EDTA, 10 μg/mL of aprotinin, and 100 μg/mL of phenylmethylsulfonyl-fluoride. Protein concentration was measured using the Bio-Rad protein assay (Bio-Rad Laboratories, Milan, Italy). Prestained molecular weight standards BenchMark were from Invitrogen (Milan). Proteins separated on the polyacrylamide gels were blotted on a nitrocellulose membrane (Protran, Whatman; Germany). The membrane was stained with Ponceau S (Sigma-Aldrich) to evaluate the success of transfer, and to locate the molecular weight markers. Free protein binding sites on the nitrocellulose were blocked with nonfat dry milk and a Tween 20/TBS solution. The membranes were washed and stained with specific primary antibodies and with secondary antisera conjugated with horseradish peroxidase diluted 1:3,000. The luminescent signal was visualized with the ECL Western blotting detection reagent kit (Amersham) and quantified by scanning with a Discover Pharmacia scanner equipped with a Sun Spark Classic Workstation. *Scion Image* version beta 4.0.3 software (Scion, Frederick, Maryland) was used to quantify signal intensity.

4.3.6 Data analysis

Each experiment was carried out 2–4 times and found to be reproducible. Human tissue samples were not pooled, therefore each sample served as its own control. Error bars are presented as standard error of the mean (SEM). We used analysis of variance to identify statistically significant differences among means. The non-parametric Spearman correlation test was used to compare breast cancer cells and control cell lines.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, our results strongly support the presence in the basal-A BRCA1-mutated cell line HCC1937 of a BCSC subpopulation bearing the phenotype $CD44^{+}CD24^{+}CD338^{bright}$, i.e. an antigenic combination different from the classical $CD44^{+}/CD24^{-/low}$.

The nature of $CD338^{low}$ and $CD338^{bright}$ cells, expressing ABCG2 at different levels, remains to be clarified. Our results suggest that $CD338^{low}$ could constitute the progeny of the most immature $CD338^{bright}$. Particularly, $CD338^{low}$ cells are likely to be the luminal progenitors blocked at this stage of development. This model would be in agreement with that proposed by Lim *et al*. who found an increased number of luminal progenitors in breast tissue from BRCA1 mutation carriers. Anyway, the nature of $CD338^{low}$ and $CD338^{bright}$ cells in HCC1937 cell line, will be elucidated by their *in vivo* behavior through the analysis of tumors originated from the sorted cell subpopulation. To conclude that CD338 might constitute a novel antigen reliable to isolate BCSCs from basal-like breast cancers, our results should be extended to other BRCA1-mutated basal-like cell lines and to human tissues from patients with basal-like breast cancers.

Considering this future perspective, concurrently to the study of breast cancer cell lines, a complementary work was carried out in the present thesis to standardize the flow-cytomertic methods to study patient-derived breast cancer cell cultures. This work will be subsequently used to isolate and study the putative $CD44^{+}CD24^{+}CD338^{bright}$ BCSC population from human breast cancer tissues. While accomplishing this flow cytometry standardization work, we have also determined whether cell cultures established from human breast cancer tissues maintain the cellular heterogeneity of primary tissues and may therefore be used for *in vitro* modeling of breast cancer subtypes. With this work we confirmed that *in vitro* culturing of breast cancer

cells reduces luminal lineage-type of cells, displaying a basal/mesenchymal phenotype. Furthermore, we demonstrated that breast cancer cell cultures preserve inter-tumor heterogeneity and express stem/progenitor markers that can be identified, quantified and categorized by flow cytometry.

VI. FLOW CYTOMETRY APPENDIX

6.1 Principles of Flow Cytometry

Flow cytometry is a technology that measures multiple physical characteristics of single particles, usually cells. The properties measured include a particle's relative size, relative granularity or internal complexit. A flow cytometer is made up of three main systems: fluidics, optics, and electronics.

• The fluidics system transports particles in a stream to the laser beam for interrogation.

• The optics system consists of lasers to illuminate the particles in the sample stream and optical filters to direct the resulting light signals to the appropriate detectors.

• The electronics system converts the detected light signals into electronic signals that can be processed by the computer.

When a sample is injected in a flow cytometer, cells are randomly distributed in the three-dimensional space. For optimal illumination by lasers, the stream transporting the cells should be positioned in the center of the laser beam. In addition, only one cell or particle should move through the laser beam at a given moment. The sample must therefore be ordered into a stream of single cells. This process is managed by the fluidics system. Essentially, the fluidics system consists of a central core through which the sample is injected, enclosed by an outer sheath that contains faster flowing fluid. When the fluidic system is activated, the flow of sheath fluid accelerates the particles and restricts them to the center of the sample core. This effect, that creates a single file of cells, is called hydrodynamic focusing. Based on principles relating to laminar flow, the sample core remains separate but coaxial within the sheath fluid and does not mix with it (Figure 6.1).

After hydrodynamic focusing, each cell passes through one or more beams of light. Light scattering or fluorescence emission (if the cell is labeled with a fluorochrome) provides information about the cell's properties.

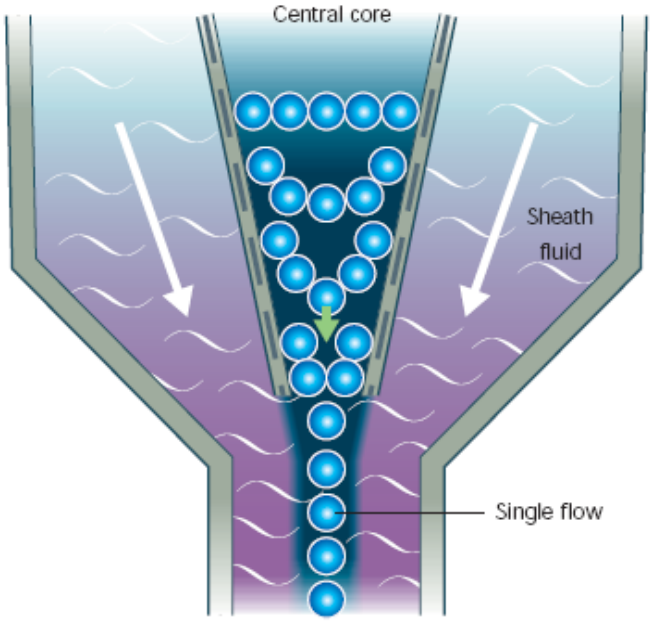


Figure 6.1: Hydrodynamic focus produces a single file of cells

The lasers are the most commonly used light sources in the modern flow cytometry. They produce a single wavelength of light at one or more frequencies. Light scattering occurs when a particle deflects incident laser light (Figure 6.2) . The extent to which this occurs depends on the physical properties of a particle. Forward-scattered light (FSC) is proportional to cell surface area or size. FSC is a measurement of mostly diffracted light and is detected just off the axis of the incident laser beam in the forward direction by a photodiode, tipically up to 20° offset from the laser beam's axis. Side-scattered light (SSC) is proportional to cell granularity or internal complexity. SSC is a measurement of mostly refracted and reflected light that occurs at any interface within the cell where there is a change in refractive index. Correlated measurements of FSC and SSC can allow for differentiation of cell types in a heterogeneous cell population just in function of their size and internal complexity.

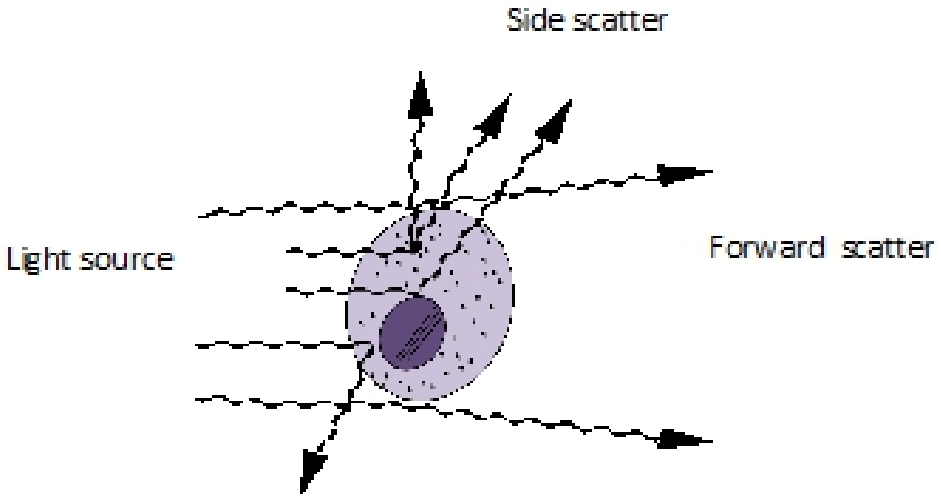


Figure 6.2: Light Scattering properties of a cell

In addition to its size and granularity, a particular cell type can be identified based on the individual antigenic surface markers by using a fluorescent dye conjugated to a monoclonal antibody. A fluorescent compound absorbs light energy over a range of wavelengths that is characteristic for that compound. This absorption of light causes an electron in the fluorescent compound to be raised to a higher energy level. The excited electron quickly decays to its ground state, emitting the excess energy as a photon of light. This transition of energy is called fluorescence. The range over which a fluorescent compound can be excited is termed its absorption spectrum. The range of emitted wavelengths for a particular compound is termed its emission spectrum. In summary, as cells intercept the light source they scatter light generating forward and side scatter light and fluorochromes are excited to a higher energy state. This energy is released as a photon of light with specific spectral properties unique to different fluorochromes antibody (Figure 6.3). So, In a mixed population of cells, different fluorochromes can be used to distinguish separate subpopulations and the amount of fluorescent signal detected is proportional to the number of fluorochrome molecules on the cell. So, the antibody staining pattern of each subpopulation, combined with FSC and SSC data, can be used to identify which cells are present in a sample and to count their relative percentages.

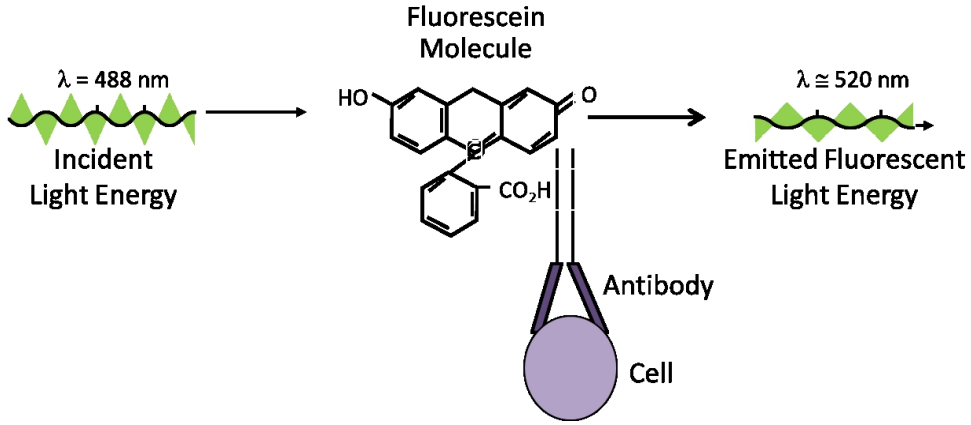


Figure 6.3: Fluorescent emission from fluorochrome-labeled cell surface molecules

The most useful fluorochromes for surface or intracellular epitope detection include single dyes and tandem dyes. Some of the single dyes, e.g. PE (figure 6.4), have been in use for the past 30 years but are now facing competition from alternatives like Alexa Fluor dyes which offer greater photostability and increased fluorescence. In a tandem dye, when the first dye is excited its energy is transferred To the second dye located in close proximity. This activates the second fluorochrome, which then produces the fluorescence emission. The process is called Fluorescence Resonance Energy Transfer (FRET). The majority of tandem dyes have been manufactured for the 488 nm laser which is found in most cytometers and they are very useful in multicolor fluorescences studies in combination with single dyes.

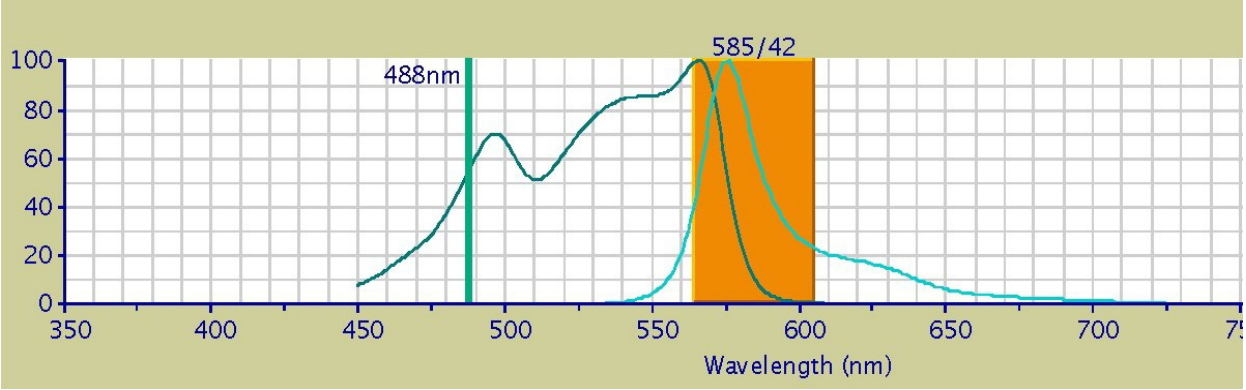


Figure 6.4: Absorpition and emission spectra of Phycoerithrin (PE) fluorochrome

The collections optics consist of a collection lens to collect light emitted from the particle–laser beam interaction and a system of optical mirrors and filters to route specified wavelengths of the collected light to designated optical detectors. Light signals generated as particles pass through the laser beam signals are converted to electronic signals (voltages) by photodetectors and then assigned a channel number on a data plot. Detectors are either photodiodes or photomultipier tubes (PMTs). Photodiodes are usually used to measure forward scatter when the signal is strong. PMTs are more sensitive instruments and are ideal for scatter and fluorescence signals. The specificity of detection is controlled by optical filters, which block certain wavelengths while transmitting (passing) others. There are three major filter types. Long pass filters allow through light above a cut-off wavelength, short-pass permit light below a cut-off wavelength and band-pass transmit light within a specified narrow range of wavelengths.

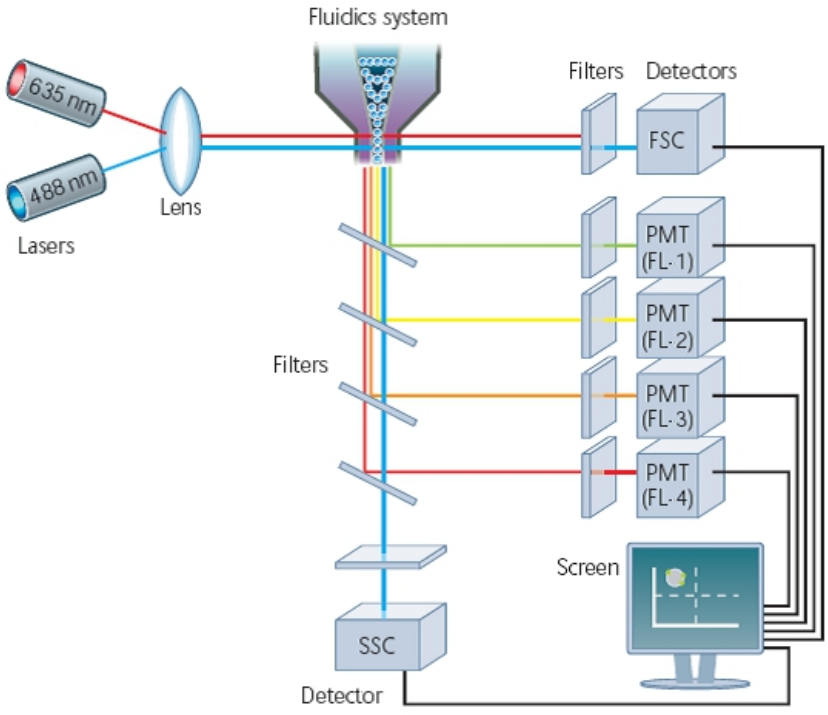


Figure 6.5: Schematic overview of optics system

6.2 Anaysis of data

Once a data file has been saved, cell populations can be displayed in several different formats. A single parameter such as SSC or FITC (fluorescence1, FL1) can be displayed as a single parameter histogram, where the horizontal axis represents the parameter's signal value in channel numbers and the vertical axis represents the number of events per channel number. Each event is placed in the channel that corresponds to its signal value. Signals with identical intensities accumulate in the same channel. Brighter signals are displayed in channels to the right of the dimmer signals. Two parameters can be displayed simultaneously in a dot plot (Fugure 6.6). One parameter is displayed on the x-axis and the other parameter is displayed on the y-axis.

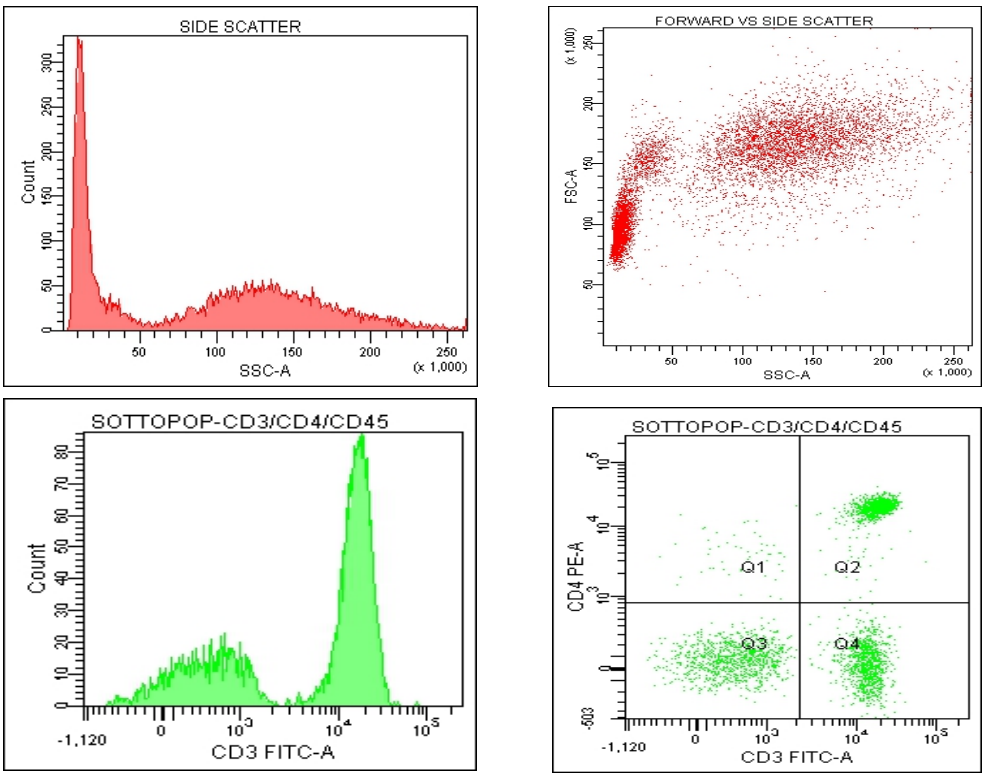


Figure 6.6: Examples of data representation: hystogram (left) and dot plot (right)

An important principle of flow cytometry data analysis is to selectively visualize the cells of interest (cells in which we want to analyze the expression of some markers) while eliminating results from unwanted particles, e.g. dead cells and debris. This procedure is called gating. A gate can be defined as a numerical or graphical boundary that can be used to define the characteristics of particles to include for further analysis. Cells are usually gated according to physical characteristics. For instance, subcellular debris and clumps can be distinguished from single cells by size, estimated by forward scatter. Also, dead cells have lower forward scatter and higher side scatter than living cells (Figure 6.7).

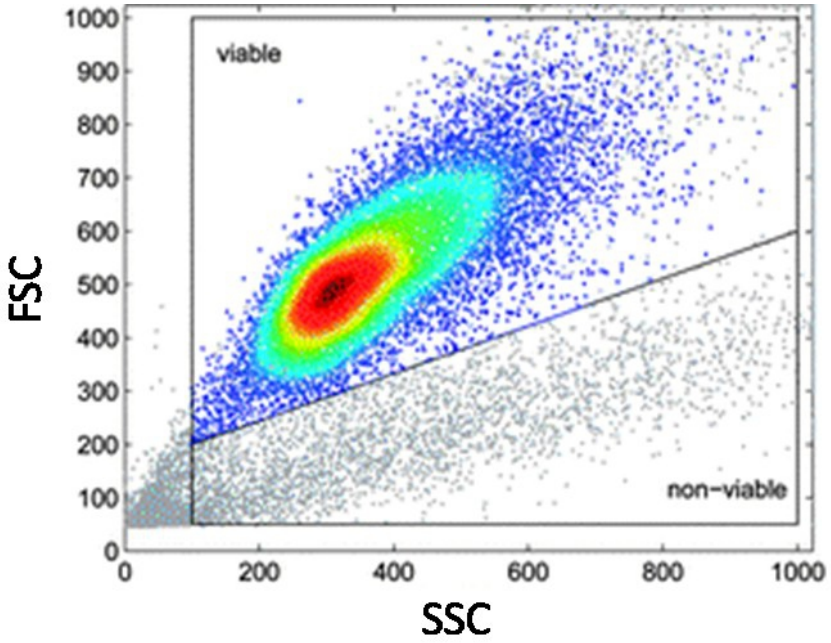


Figure 6.7: Exclusion of dead cells according to physical parameters forward and side scatter

After exclusion of debris, clumps and dead cells, it's important to remove doublets from the subset of cells target of the analysis, because sometimes doublets can lead to false percentages. One of the most used method to discriminate single cells from doublets it's the dot plot forward scatter-area (FSC-A) *vs* forward scatter-height (FSC-H) in which every events that are out of the diagonal are considered doublets because a doublet has the same

height value of a single cell but an higher area value. Cells passing through cytometer generate voltage pulses of characteristic shape, gaussian for round shaped cells (Figure 6.8). The ratio of area under the curve of the pulse to its width (x height) is different for single cells (Figure 6.8, upper left panel) and doublets (Figure 6.8, upper right panel) due to the increase in time that light is scattered for when a doublet passes through compared to a singlet. In other words, doublets will have greater pulse width than a single cell, as they take longer to pass through the laser beam.

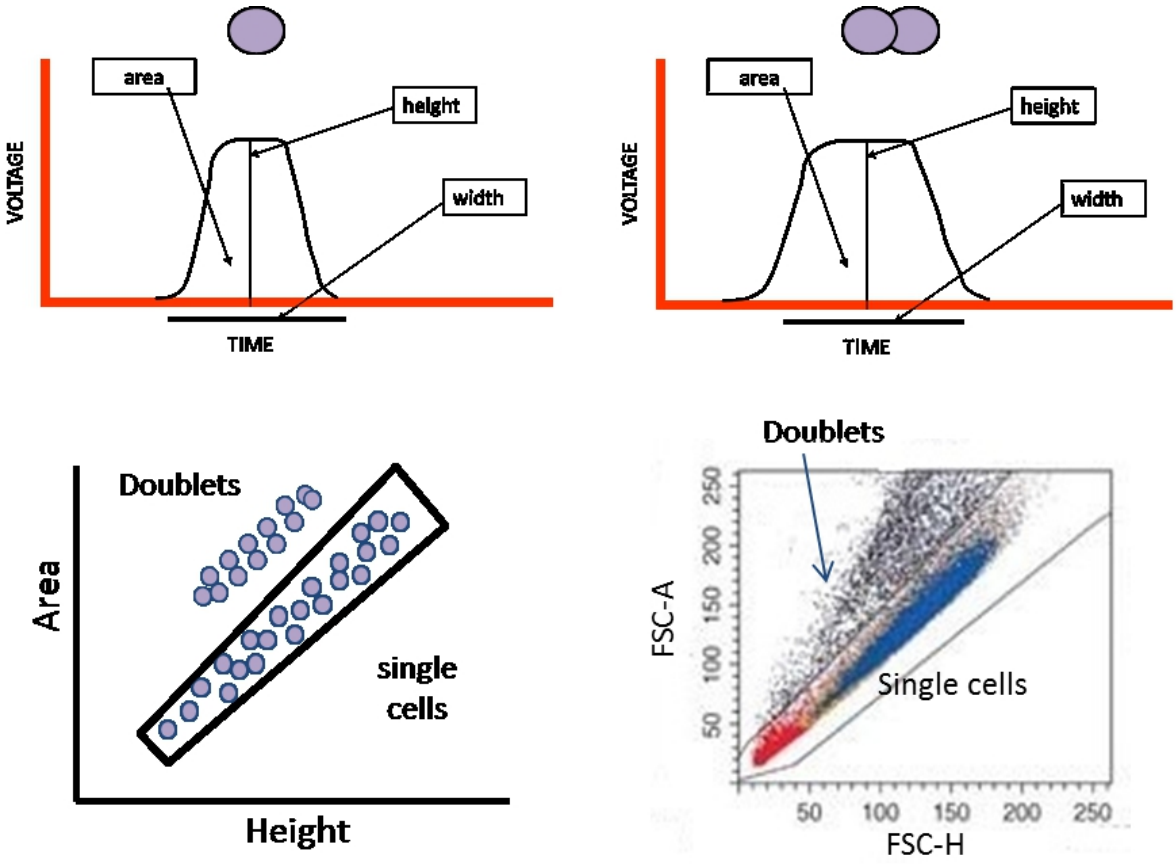


Figure 6.8: Exclusion of doublets according to physical parameters forward-area (FSC-A) and forward-height scatter (FSC-H).

Once debris, dead cells and doublets have been excluded with gates on phisical parameter dot plots, the gated cell subset can be further analyzed to evaluate the expression of cell surface markers using fluorescent signals emitted by the fluorochrome-conjugated antidodies bound to the cell membrane. Generally, the expression of surface markers can be evaluated using two different approaches. The most common consists to define the percentage of cells which are positive for the marker indicating how many cells express (positive) and not express (negative) that marker in the target cell population. The percentage of positive cells is established by using a negative control. Generally, in flow cytometry analysis two kinds of negative controls are important: 1) the isotypic control, that allows to show aspecific bindings between the Igs and the cellular membrane; 2) the unstained cells, that allow to show the autofluorescence of cells. More particularly, the isotypic control is essential only when the analyzed cells express the receptor for the Fc fragment of Ig. Moreover, in multicolor flow cytometry analysis, when the expression of a high number of antigens is analyzed at the same time, it is usual to find some antigens expressed (positive) and others that are not expressed (negative). This allows to have many internal negative controls that can be considered "isotypic". In other words, the negative antigens can be used as controls for the positive ones. The percentage is very useful when the sample analyzed is heterogeneous including different types of cells that can or cannot express that marker. In addition to the percentage, the expression of surface markers can be also evaluated by fluorescence intensity that is proportional to the number of antibodies bound to the membrane and so to the number of recognized molecules present on the cell membrane. The fluorescence intensity gives an information different from that given by the percentage: it indicates the level of expression of a molecule on a particular cell type. So, the percentage can be used to identify different cell types in the sample analyzed and the fluorescence intensity to define the expression level of many molecules on each cell type.

6.3 Cell Sorting

In most applications, after a particle exits the laser beam, it is sent to waste. Sorting allows to capture and collect cells of interest for further analysis. Once collected, the cells can be analyzed microscopically, biochemically, or functionally. To sort particles or cells, the cytometer first needs to identify the cells of interest, then separate out the individual cells. Once the population of interest has been identified on a data acquisition plot, a gate is drawn around that population. This gate is then loaded into the cytometer's software as the "sort gate" that identifies cells of interest to be sorted out of the stream. In particular, in order to sort the fluid stream containing cells must be precisely, and reproducibly over a substantial time frame, broken into drops. To this aim, during sorting drop drive energy is applied to the stream through a vibrating nozzle to break it into highly uniform droplets. In particular, the nozzle is vibrated by a transducer which converts electrical energy into mechanical energy Droplets detach from the stream a few millimeters downstream from the nozzle. When a particle is detected that meets the predefined sorting criteria, an electrical charge is applied to the stream just as the droplet containing that particle breaks off from the stream. Once broken off from the stream, the drople still retains its charge. The charged droplet passes by two strongly charged deflection plates. Electrostatic attraction and repulsion cause each charged droplet to be deflected to the left or right, depending on the droplet's charge polarity. Uncharged droplets are not affected by the electric field and pass down the center to the waste aspirator (Figure 6.9).

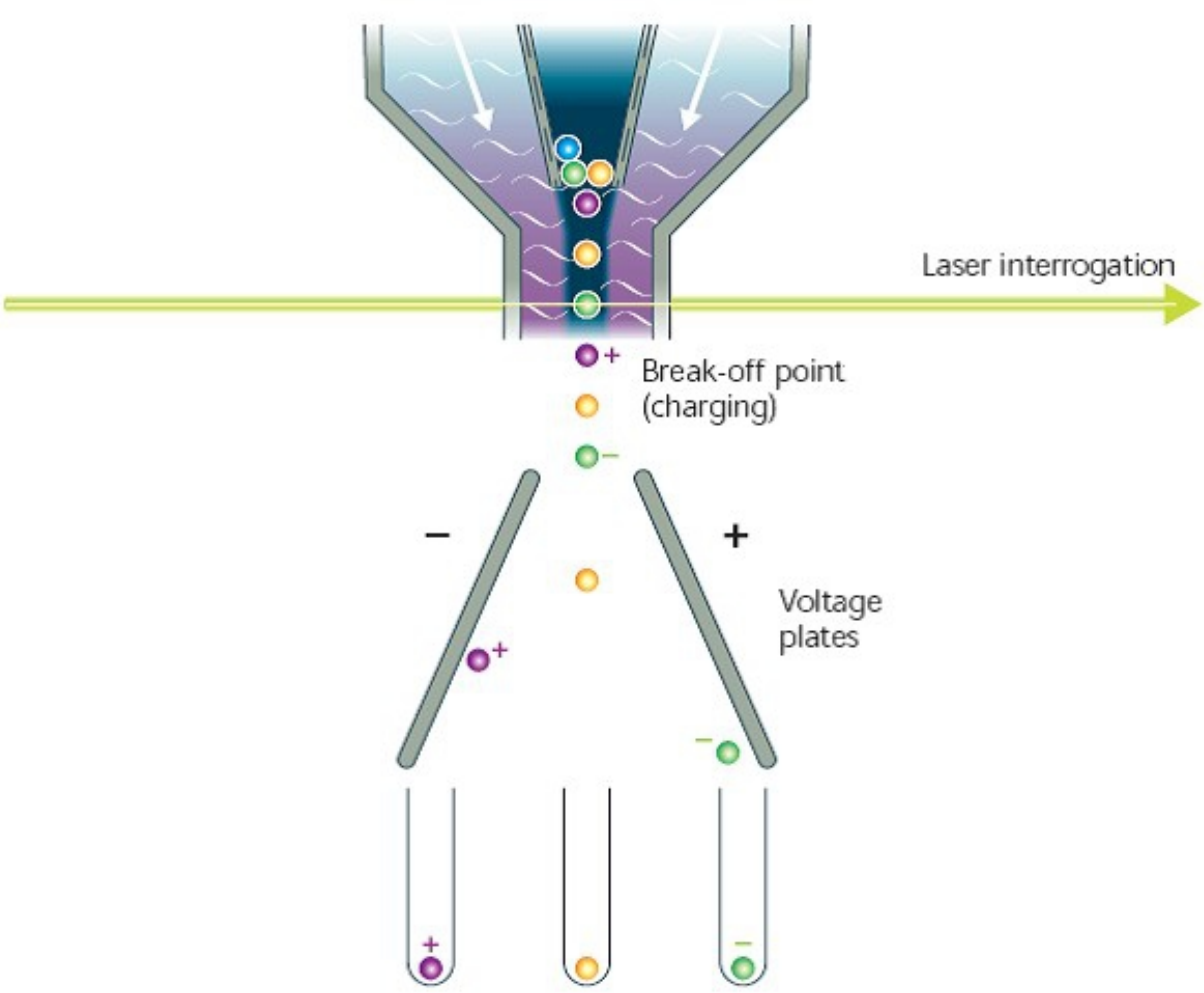


Figure 6.9: Schematic representation of cell sorting

6.4 Cytometers used in the study

During this study different types of cytometers have been used. The most part of analysis and cell sorting were performed with FACSAria I cell sorter (Becton Dickinson, Franklin Lakes, NJ). The BD FACSAria instrument uses low-powered air-cooled and solid-state lasers that do not have special power and cooling requirements. The instrument is equipped with three lasers: 488-nm blue laser, 633-nm, red laser and 407-nm violet laser. Moreover, the 375nm near UV optional laser has been installed. The 375-nm laser and the violet laser share the same beam spot at the point of interception within the cuvette, so I had to select to run either the 375- nm laser, or the violet laser.

To analyze the expression of CD338 and to sort out the different CD338 subpopulations from HCC1937 cell line, a PE conjugated antibody has been used in the present study. PE fluorescence of CD338 was determined by a 488nm excitation line (blue laser) and detected by 585/42 nm filters (Table 13). The 488 nm laser allowed to discriminate CD338 positive cells from CD338 negative cells.

Table 13: Default setup for detector array in FACSAria I cell sorter. In red are underlined the laser and the filter used to respectively excite and detect the PE fluorochrome conjugated to the CD338 antibody.

Detector Array (Laser)	PMT	LP Mirror	BP Filter	Intended Dye
Octagon (488-nm blue laser)	A	735	780/60	PE-Cy7
	B	655	695/40 675/20	PerCP-Cy5.5 or PI PerCP alone
	C	595	610/20	PE-Texas Red
	D	556	575/26 585/42	PE or PI Alternative for PE/PI when not using PE-Texas Red
	E	502	530/30	FITC
	F	—	488/10	Side scatter (SSC)
Trigon (633-nm red laser)	A	735	780/60	APC-Cy7
	B	—	660/20	APC
Trigon (407-nm violet laser)	A	502	530/30	Alexa Fluor 430, Hoeschst, DAPI
	B	—	450/40	Cascade blue, Pacific blue, Alexa Fluor 405

Later, I had the opportunity to sort the different CD338 subpopulations with FACSAria III cell sorter that is equipped with five lasers: 488nm, 561nm, 640nm, 405nm, 355nm. The 561nm yellow-green laser is a better laser than 488 blue laser to excite PE. It allows to better discriminate the different fluorescence intensities emitted by PE fluorochrome. So, I performed some cell sorting of CD338 negative and positive cells by using the yellow-green laser, but the purity of cell sorting was very low (about 20%) (Figure 6.10). I carried out many cytometric tests to understand if such a low purity was due to the cells, the antibody, the fluorochrome or the cyometer and finally I found that there was a problem just with the yellow-green laser of FACSARIA III cell sorter. So, despite the yellow-green laser is the best laser to excite PE, unfortunately I could not use it with FACSAria III cell sorter and I had to come back to use the blue laser of FACSAria I to sort CD338 populations out of HCC1937 cell line.

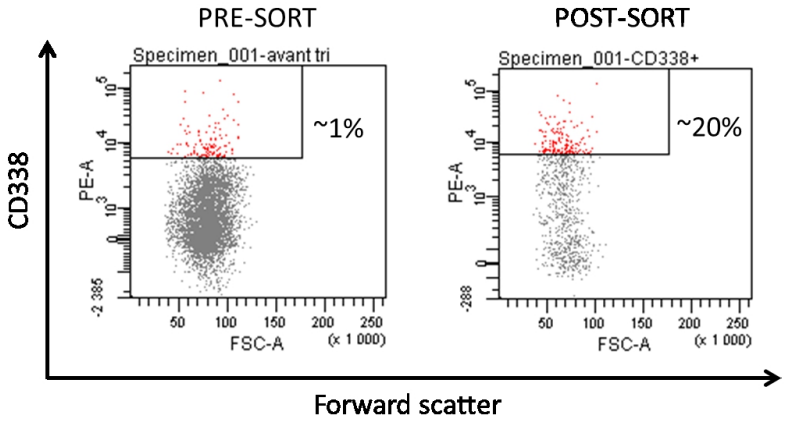


Figure 6.10: Purity of cell sorting performed with 561 nm green laser of FACSAria III

However, I could use the same 561nm yellow-green laser of another well-working cytometer, LSR II. Unfortunately, LSR II is not a cell sorter but just an analyzer, so I could not use it to sort HCC1937 cells but just to analyze them. In particular I used it: 1) to discriminate, among the CD338 negative cells, the real negative cells from a subpopulation of cells expressing CD338 at low level, named $CD338^{+/low}$ (Figure 6.11, upper right panel); 2) to check the purity of cell sorting performed with FACSAria I; 3) to follow the expression of CD338 in cultured sorted cells. In particular, with LSR II cytometer I could analyze the expression of CD338 with both lasers, 488nm and 561nm, at the same time. The comparison of data obtained with the two lasers, allowed me to have a clear view on the real expression of CD338 in HCC1937 cell line and on the nature of sorted cell subpopulations. Moreover this comparison allowed me to try to sort the three CD338 subpopulations, $CD338^{+/bright}$ $CD338^{+/low}$ and $CD338^{neg}$, with the blue laser of FACSAria I, on the base of the two images given by the blue and the yellow-green lasers of LSR II cytometer (Figure 6.11).

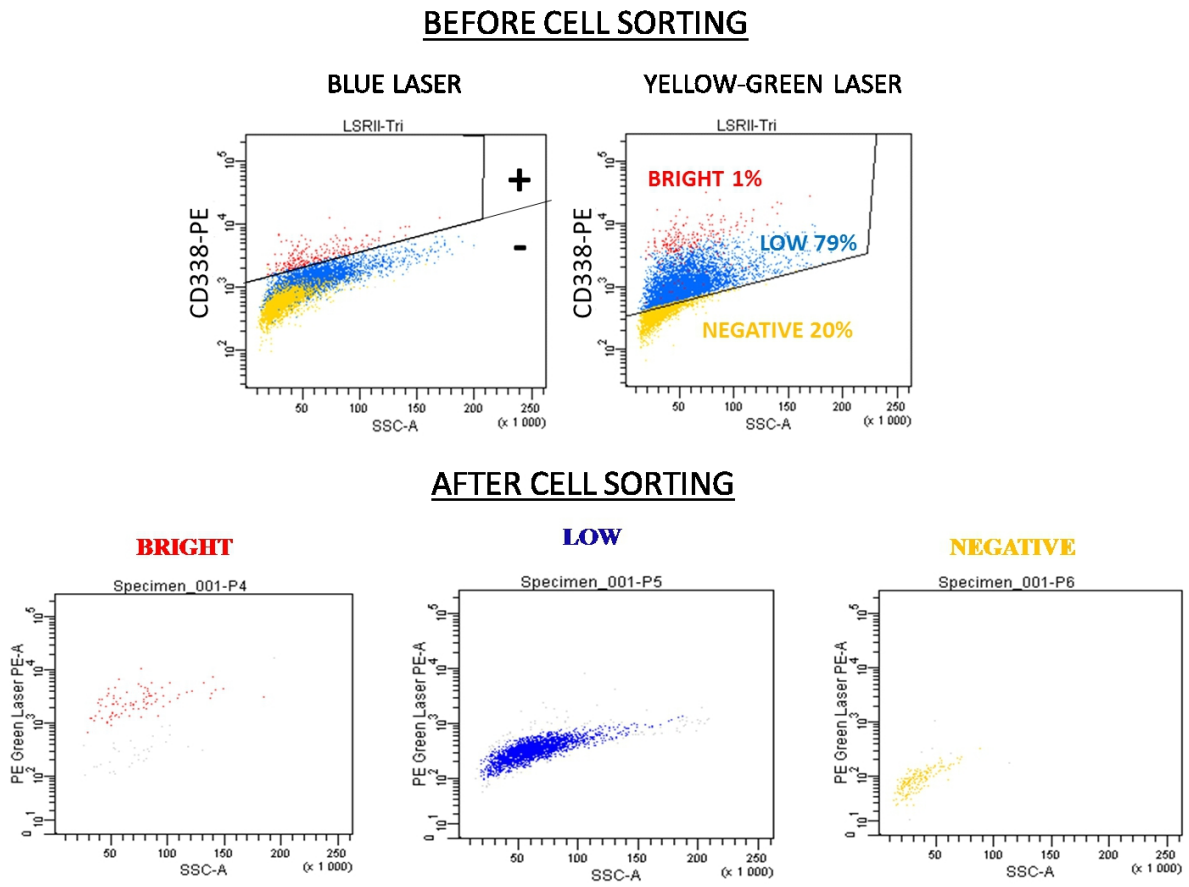


Figure 6.11. Expression of CD338 in HCC1937 cell line and cell sorting of three cell subsets. Upper panels: FACSAria I Cytometer is equipped with the blue laser that is able to discriminate just CD338 positive (upper left panel, red events) and CD338 negative (upper left panel, blue and yellow events). LSR II Cytometer is equipperd also with the yellow-green laser that is able to discriminate three different cell subsets: bright, low and negative cells (upper right panel). Blue events, that seemed to be CD338 negative cells according to the blue laser, actually are CD338 positive (as shown by the yellow laser), although expressing CD338 at lower level than $CD338^{bright}$ cells. Lower panels: example of sorting of the three subsets performed with FACAria I sorter on the base of the images given by the two lasers of LSR II analyzer.

VII. REFERENCES

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA 100, 3983–3988.

Anbazhagan R, Osin PP, Bartkova J, Nathan B, Lane EB, Gusterson BA. (1998). The development of epithelial phenotypes in the human fetal and infant breast. J Pathol 184(2), 197-206.

Apati A, Orban TI, Varga N, Nemeth A, Schamberger A, Krizsik V, et al. (2008). High level functional expression of the ABCG2 multidrug transporter in undifferentiated human embryonic stem cells. Biochimica et Biophysica Acta 1778 (12), 2700–2709.

Babiker AA, Nilsson B, Ronquist G, Carlsson L, Ekdahl KN. (2005). Transfer of functional prostasomal CD59 of metastatic prostatic cancer cell origin protects cells against complement attack. Prostate 62(2), 105-114.

Bachelard-Cascales E, Chapellier M, Delay E, Pochon G, et al. (2010). The CD10 enzyme is a key player to identify and regulate human mammary stem cells. Stem Cells 28(6), 1081-8.

Badve S, Dabbs DJ, Schnitt SJ, Baehner FL, Decker T, Euseb V, et al. (2011). Basal-like and triple-negative breast cancers:a critical review with an emphasis on the implications for pathologists and oncologists. Modern Pathology 24, 157–167.

Bapat SA, Mali AM, Koppikar CB, Kurrey NK. (2005). Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. Cancer Research 65, 3025–3029.

Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. (2003). Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone

for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol. 21(10), 1973–1979.

Bartlett JM, Brookes CL, Robson T, van de Velde CJ, Billingham LJ, Campbell FM, Grant M et al. (2011). Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: a prospectively powered pathology study in the Tamoxifen and Exemestane Adjuvant Multinational trial. J Clin Oncol. 29(12), 1531–1538.

Bartek J, Taylor-Papadimitriou J, Miller N, Millis R. (1985) Patterns of expression of keratin 19 as detected with monoclonal antibodies in human breast tissues and tumours. Int J Cancer 36(3), 299-306.

Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, et al. (2006). Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45, 1033–1040.

Bertos NR, Park M. (2011). Breast cancer - one term, many entities? J Clin Invest. 121(10):3789-96.

Bhattacharya S, Das A, Mallya K, Ahmad I. (2007). Maintenance of retinal stem cells by Abcg2 is regulated by notch signaling. Journal of Cell Science 120 (Pt 15), 2652–2662.

Blick T, Widodo E, Hugo H, Waltham M, Lenburg ME, Neve RM, Thompson EW. (2008). Epithelial mesenchymal transition traits in human breast cancer cell lines. Clin Exp Metastasis 25(6), 629-642.

Boman BM, Wicha MS, Fields JZ, Runquist OA. (2007). Symmetric Division of Cancer Stem Cells – a Key Mechanism in Tumor Growth that should be Targeted in Future Therapeutic Approaches. Clin Pharmacol Ther 81, 893-898.

Bochar DA, Wang L, Beniya H, et al. (2000). BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. Cell 102(2), 257–65.

Bonnet D, Dick JE. (1994). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature 17;367(6464), 645-8.

Buckley NE, Mullan PB. (2012). BRCA1--conductor of the breast stem cell orchestra: the role of BRCA1 in mammary gland development and identification of cell of origin of BRCA1 mutant breast cancer. Stem Cell Rev.8(3):982-93.

Cai J, Cheng A, Luo Y, et al. (2004). Membrane properties of rat embryonic multipotent neural stem cells. J Neurochem 88, 212–26.

Carey L, Winer E, Viale G, Cameron D, Gianni L. (2010). Triple-negative breast cancer: disease entity or title of convenience? Nat Rev Clin Oncol. 7(12):683-92.

Cariati M, Naderi A, Brown JP, Smalley MJ, Pinder SE, Caldas C, Purushotham AD. (2008). Alpha-6 integrin is necessary for the tumourigenicity of a stem cell-like subpopulation within the MCF7 breast cancer cell line. Int J Cancer 122(2), 298-304.

Carter WG, Kaur P, Gil SG, Gahr PJ, Wayner EA. (1990). Distinct functions for integrins α3β1 in focal adhesions and α6β4/bullous pemphigoid antigen in a new stable anchoring contact (SAC) of keratinocytes: relation to hemidesmosomes. J Cell Biol 111, 3141–3154.

Chamberlain G, Fox J, Ashton B, Middleton J. (2007). Concise review: mesenchymal stem cells: their phenotype, differentiation capacity, immunological features, and potential for homing. Stem Cells 25(11), 2739-2749.

Charafe-Jauffret E, Ginestier C, Birnbaum D. (2009). Breast cancer stem cells: tools and models to rely on. BMC Cancer 9, 202.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, et al. (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10, 529–541.

Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, Visvader J, Weissman IL, Wahl GM. (2006). Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Res 66(19), 9339-9344.

Clayton H, Titley I, Vivanco M. (2004). Growth and differentiation of progenitor/stem cells derived from the human mammary gland. Exp Cell Res 297, 444-460.

Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. (2005). Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res 65, 10946–10951.

Connolly J, Kempson R, LiVolsi V, Page D, Patchefsky A, Silverberg S. (2004). Recommendations for the reporting of breast carcinoma. Association of Directors of Anatomic and Surgical Pathology.

Dent R, Trudeau M, Pritchard KI et al. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. Clin. Cancer Res. 13, 4429–4434.

Desuzinges-Mandon E, Arnaud O, Martinez L, Huché F, Di Pietro A, Falson P. (2010). ABCG2 transports and transfers heme to albumin through its large extracellular J Biol Chem. 285(43), 33123-33.

Donnenberg VS, Donnenberg AD, Zimmerlin L, Landreneau RJ, Bhargava R, Wetzel RA, Basse P, Brufsky AM. (2010). Localization of CD44 and CD90 positive cells to the invasive front of breast tumors. Cytometry B Clin Cytom 78(5), 287-301.

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 17, 1253-1270.

Eirew P, Stingl J, Raouf A, Turashvili G, Aparicio S, Emerman JT, et al. (2008). A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. Nat Med 14(12), 1384-9.

Esteban JM, Felder B, Ahn C, Simpson JF, Battifora H, Shively JE. (1994). Prognostic relevance of carcinoembryonic antigen and estrogen receptor status in breast cancer patients. Cancer 74(5), 1575-1583.

Fang D, Nguyen TK, Leishear K, Finko R, Kulp AN, Hotz S, Van Belle PA, Xu X, Elder DE, Herlyn M. (2005). A tumorigenic subpopulation with stem cell properties in melanomas. Cancer Res. 65, 9328–9337.

Feuerhake F, Sigg W, Hofter EA, Dimpfl T & Welsch, U (2000). Immunohistochemical analysis of Bcl-2 and Bax expression in relation to cell turnover and epithelial differentiation markers in the nonlactating human mammary gland epithelium. Cell Tissue Res. 299, 47–58.

Fillmore CM, Kuperwasser C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res. 10(2), R25.

Foulkes, W. D. (2004). BRCA1 functions as a breast stem cell regulator. Journal of Medical Genetics 41(1), 1–5.

Foulkes WD, Brunet JS, Stefansson IM et al. (2004). The prognostic implication of the basal-like (cyclin E high / p27 low / p53+ / glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. Cancer Res. 64, 830–835.

Fukuda K, Saikawa Y, Ohashi M, Kumagai K, Kitajima M, Okano H, Matsuzaki Y, Kitagawa Y. (2009). Tumor initiating potential of side population cells in human gastric cancer. Int J Oncol. 34, 1201–1207.

Gautam K. Malhotra, Xiangshan Zhao, Hamid Band and Vimla Band. (2010). Histological, molecular and functional subtypes of breast cancers. Cancer Biology & Therapy 10:10, 955-960.

Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 1(5), 555-567.

Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med. 183(4):1797-806.

Graham FL, van der Eb AJ. (1973). Transformation of rat cells by DNA of human adenovirus 5. Virology. 54(2):536-9.

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

Guelstein V, Tchypysheva T, Ermilova V, Litvinova L, Troyanovsky S, Bannikov G. (1988). Monoclonal antibody mapping of keratins 8 and 17 and of vimentin in normal human mammary gland, benign tumors, dysplasias and breast cancer. Int J Cancer 42, 147-153.

Gugliotta P, Sapino A, Macri L, Skalli O, Gabbiani G, Bussolati G. (1998). Specific demonstration of myoepithelial cells by anti-alpha smooth muscle actin antibody. J Histochem Cytochem 36, 659-663.

Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? (2009). Nat Med. 15(9), 1010-2.

Gusterson BA, Warburton MJ, Mitchell D, Ellison M, Neville AM, Rudland PS. (1982). Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. Cancer Res 42(11), 4763-4770.

Gusterson B, Monaghan P, Mahendran R, Ellis J, O'Hare M. (1986). Identification of myoepithelial cells in human and rat breasts by anti-common acute lymphoblastic leukemia antigen antibody A12. J Natl Cancer Inst 77, 343-349.

Gusterson BA, Ross DT, Heath VJ, Stein T. (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. Breast Cancer Res. 7, 143–148.

Haffty BG, Yang Q, Reiss M et al. (2006). Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. J. Clin. Oncol. 24, 5652–5657.

Harris LN, Broadwater G, Lin NU et al. (2006). Molecular subtypes of breast cancer in relation to paclitaxel response and outcomes in women with metastatic disease: results from CALGB 9342. Breast Cancer Res. 8, R66.

Harrison DE & Lerner CP. (1991). Most primitive hematopoietic stem cells are stimulated to cycle rapidlyafter treatment with 5-fluorouracil. Blood 78, 1237–1240.

Havens AM, Jung Y, Sun YX, Wang J, Shah RB, Bühring HJ, Pienta KJ, Taichman RS. (2006). The role of sialomucin CD164 (MGC-24v or endolyn) in prostate cancer metastasis. BMC Cancer 6, 195.

Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, et al. (2009). Characterization of a naturally occurring breast cancer subset enriched in epithelial- to-mesenchymal transition and stem cell characteristics. Cancer Res. 69(10), 4116–4124.

Henry LA, Johnson DA, Sarrió D, Lee S, Quinlan PR, Crook T, Thompson AM, Reis-Filho JS, Isacke CM. (2011). Endoglin expression in breast tumor cells suppresses invasion and metastasis and correlates with improved clinical outcome. Oncogene 30(9), 1046-1058.

Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, Bruns CJ, Heeschen C. (2007). Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell. 1, 313–323.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 8, 76.

Hirschmann-Jax C, Foster AE, Wulf GG, Nuchtern JG, Jax TW, Gobel U, et al. (2004). A distinct "side population" of cells with high drug efflux capacity in human tumor cells. Proc Natl Acad Sci USA 101, 14228-14233.

Ho MM, Ng AV, Lam S, Hung JY. (2007). Side population in human lung cancer cell lines and tumors is enriched with stem-like cancer cells. Cancer Res. 67, 4827–4833.

Honeth G, Bendahl PO, Ringnér M, Saal LH, Gruvberger-Saal SK, Lövgren K, Grabau D, Fernö M, Borg A, Hegardt C. (2008). The CD44+/CD24- phenotype is enriched in basal-like breast tumors. Breast Cancer Res. 10(5), 110.

Hovey RC, McFadden TB, Akers RM. (1999). Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. J Mammary Gland Biol Neoplasia 4, 53-68.

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7, 96.

Huang CC, Tu SH, Lien HH, Jeng JY, Liu JS, Huang CS, Wu YY, Liu CY, Lai LC, Chuang EY. (2012). Prediction consistency and clinical presentations of breast cancer molecular subtypes for Han Chinese populationJ Transl Med. Epub 2012 Sep 19.

Hwang-Verslues WW, Kuo WH, Chang PH, Pan CC, et al. (2009). Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers. PLoS One 4(12), e8377.

Ihnen M, Kilic E, Köhler N, Löning T, Witzel I, Hagel C, Höller S, Kersten JF, Müller V, Jänicke F et al. (2011). Protein expression analysis of ALCAM and CEACAM6 in breast cancer metastases reveals significantly increased ALCAM expression in metastases of the skin. J Clin Pathol 64(2), 146-152.

Ikeda J, Morii E, Liu Y, Qiu Y, Nakamichi N, Jokoji R, Miyoshi Y, Noguchi S, Aozasa K. (2008). Prognostic significance of CD55 expression in breast cancer. Clin Cancer Res 14(15), 4780-4786.

Ilantzis C, DeMarte L, Screaton RA, Stanners CP. (2002). Deregulated expression of the human tumor marker CEA and CEA family member CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation. Neoplasia 4(2), 151-163.

Islam MO, Kanemura Y, Tajria J, Mori H, Kobayashi S, Hara M, Yamasaki et al. (2005). Functional expression of ABCG2 transporter in human neural stem/progenitor cells. Neuroscience Research 52 (1), 75–82.

Iwaya K, Ogawa H, Izumi M, Kuroda M, Mukai K. (2002). Stromal expression of CD10 in invasive breast carcinoma: a new predictor of clinical outcome. Virchows Arch 440(6), 589-593.

Jordan VC, Brodie AM. (2007). Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. Steroids 72(1), 7–25.

Junttila TT, Akita RW, Parsons K, Fields C, Lewis Phillips GD, Friedman LS, Sampath D, Sliwkowski MX. et al. (2009). Ligand-independent HER2/HER3/PI3K complex is disrupted by trastuzumab and is effectively inhibited by the PI3K inhibitor GDC-0941. Cancer Cell 15(5), 429–440.

Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, et al. (2009). Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 4(7), e6146.

Kawasaki BT, Mistree T, Hurt EM, Kalathur M, Farrar WL. (2007). Co-expression of the toleragenic glycoprotein, CD200, with markers for cancer stem cells. Biochem Biophys Res Commun 364(4), 778-782.

Keller PJ, Lin A, Arendt LM, Klebba I, Jones AD, Rudnick JA, Dimeo TA, et al. (2010). Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res 12(5), R87.

Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, et al. (2011). Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A 109(8), 2772-7.

Kordon EC & Smith GH. (1998). An entire functional mammary gland may comprise the progeny from a single cell. Development 125, 1921–1930.

Krohn A, Song YH, Muehlberg F, Droll L, Beckmann C, et al. (2009). CXCR4 receptor positive spheroid forming cells are responsible for tumor invasion in vitro. Cancer Lett 280(1), 65-71.

Kulasingam V, Zheng Y, Soosaipillai A, Leon AE, Gion M, Diamandis EP. (2009). Activated leukocyte cell adhesion molecule: a novel biomarker for breast cancer. Int J Cancer 125(1), 9-14.

Lacroix M, Leclercq G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat 83, 249–289.

Lakhani SR, Reis-Filho JS, Fulford L et al. (2005). Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin. Cancer Res. 11, 5175–5180.

Lakhani SR, O'Hare MJ. (2001). The mammary myoepithelial cell –Cinderella or ugly sister? Breast Cancer Res. 3, 1–4.

Langsenlehner U, Renner W, Yazdani-Biuki B, Eder T, Wascher TC, Paulweber B, et al. (2006). Integrin alpha-2 and beta-3 gene polymorphisms and breast cancer risk. Breast Cancer Res Treat 97(1), 67-72.

Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA, Dick JE. (2004). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 432(7015), 396-401.

Lasa A, Serrano E, Carricondo M, Carnicer MJ, Brunet S, Badell I, Sierra J, Aventín A, Nomdedéu JF. (2008). High expression of CEACAM6 and CEACAM8 mRNA in acute lymphoblastic leukemias. Ann Hematol 87(3), 205-211.

Latza U, Niedobitek G, Schwarting R, Nekarda H, Stein H: Ber- EP4. (1990). A new monoclonal antibody which distinguishes epithelia from mesothelial. J Clin Pathol 43, 213-219.

Lechner A, Leech CA, Abraham EJ, Nolan AL, Habener JF. (2002). Nestin-positive progenitor cells derived from adult human pancreatic islets of Langerhans contain side population (SP) cells defined by expression of the ABCG2 (BCRP1) ATP-binding cassette transporter. Biochem Biophys Res Commun 293, 670–4.

Li CI, Uribe DJ, Daling JR. (2005). Clinical characteristics of different histologic types of breast cancer. Br J Cancer 93, 1046-52.

Lim E, Vaillant F, Wu D, et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nature Medicine, 15 (8), 907–13.

Lindeman GJ, Visvader JE. (2010). Insights into the cell of origin in breast cancer and breast cancer stem cells. Asia Pac J Clin Oncol. 6(2),89-97.

Liu S, Ginestier C, Charafe-Jauffret E, et al. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. Proceedings of the National Academy of Sciences of the United States of America, 105(5), 1680–5.

Mahler-Araujo B, Savage K, Parry S, Reis-Filho JS. (2008). Reduction of E-cadherin expression is associated with non-lobular breast carcinomas of basal-like and triple negative phenotype. J Clin Pathol. 61(5), 615-20.

Malhotra GK, Zhao X, Band H, Band V. (2010). Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther. 10(10):955-60.

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 133(4), 704-15.

Manna PP, Frazier WA. (2004). CD47 mediates killing of breast tumor cells via Gi-dependent inhibition of protein kinase A. Cancer Res 64(3), 1026-1036.

Meyer MJ, Fleming JM, Ali MA, Pesesky MW, Ginsburg E, Vonderhaar BK. (2009). Dynamic regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell lines. Breast Cancer Res 11, R82.

Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, et al. (2007). Lung metastasis genes couple breast tumor size and metastatic spread. Proc Natl Acad Sci USA 104, 6740-6745.

Molyneux, G., Geyer, F. C., Magnay, F. A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell, 7(3), 403–17.

Montanaro F, Liadaki K, Schienda J, Flint A, Gussoni E, Kunkel LM. (2004). Demystifying SP cell purification: viability, yield, and phenotype are defined by isolation parameters. Exp Cell Res 298, 144-154.

Morel AP, Lièvre M, Thomas C, Hinkal G, Ansieau S, Puisieux A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS One 3(8), e2888.

Morris GJ, Naidu S, Topham AK et al. (2007). Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and end results database. Cancer 110, 876–884.

Morrison BJ, Schmidt CW, Lakhani SR, Reynolds BA, Lopez JA. (2008). Breast cancer stem cells: implications for therapy of breast cancer. Breast Cancer Res. 10(4), 210.

Morrison SJ, Kimble J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441, 1068-1074.

Mullan, PB, Hosey, AM, Buckley, N, et al. (2005). The 2,5 oligoadenylate synthetase/RNaseL pathway is a novel effector of BRCA1- and interferon-gamma-mediated apoptosis. Oncogene 24(35), 5492-501.

Nagle RB, Böcker W, Davis JR, Heid HW, Kaufmann M, Lucas DO, Jarasch ED. (1986). Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. J Histochem Cytochem 34(7), 869-881.

Nardone A, Corvigno S, Brescia A, D'Andrea D, Limite G, Veneziani BM (2011). Long-term cultures of stem/progenitor cells from lobular and ductal breast carcinomas under non-adherent conditions. Cytotechnology 63(1), 67-80.

Nardone A, Cavaliere C, Corvigno S, Limite G, De Placido S, Veneziani BM. (2009). A banking strategy toward customized therapy in breast cancer. Cell Tissue Bank 10(4), 301-308.

Narod, SA, and Foulkes, WD. (2004). BRCA1 and BRCA2: 1994 and beyond. Nat. Rev. Cancer 4, 665–676.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F et al. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10(6), 515-527.

Niehans GA, Cherwitz DL, Staley NA, Knapp DJ, Dalmasso AP. (1996). Human carcinomas variably express the complement inhibitory proteins CD46 (membrane cofactor protein), CD55 (decay-accelerating factor), and CD59 (protectin). Am J Pathol 149(1), 129-142.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, et al. (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 10(16), 5367–5374.

Nishida H, Yamazaki H, Yamada T, Iwata S, Dang NH, Inukai T, Sugita K, Ikeda Y, Morimoto C. (2009). CD9 correlates with cancer stem cell potentials in human B-acute lymphoblastic leukemia cells. Biochem Biophys Res Commun 382(1), 57-62.

Nishimura R, Arima N. (2008). Is triple negative a prognostic factor in breast cancer? Breast Cancer 5(4), 303–308.

O'Brien CA, Pollett A, Gallinger S, Dick JE. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 445, 106–110.

Ozvegy-Laczka, C, Varady, G, Koblos, G, Ujhelly, O, Cervenak, J, Schuetz, JD, et al. (2005). Function-dependent conformational changes of the ABCG2 multidrug transporter modify its interaction with a monoclonal antibody on the cell surface. J. Biol. Chem. 280, 4219–4227.

Ozvegy-Laczka C, Laczkó R, Hegedus C, Litman T, Várady G, Goda K, (2008). Interaction with the 5D3 monoclonal antibody is regulated by intramolecular rearrangements but not by covalent dimer formation of the human ABCG2 multidrug transporter. J Biol Chem. 283(38), 26059-70.

Paik S, Kim C, Wolmark N. (2008). HER2 status and benefit from adjuvant trastuzumab in breast cancer. N Engl J Med 358(13), 1409-1411.

Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K. (2010). Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. Clin Cancer Res. 16(3), 876-87.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 27(8), 1160–1167.

Patrawala L, Calhoun T, Schneider-Broussard R, Zhou J, Claypool K, Tang DG. (2005). Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic Cancer Res. 65(14), 6207-19.

Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, Bernard L, Viale G, Pelicci PG, Di Fiore PP. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell 140(1), 62-73.

Peddi PF, Ellis MJ, Ma C. (2012). Molecular basis of triple negative breast cancer and implications for therapy. Int J Breast Cancer.

Peinado H, Olmeda D, Cano A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer 7(6), 415-428.

Penland SK, Keku TO, Torrice C et al. (2007). RNA expression analysis of formalin-fixed paraffin-embedded tumors. Lab. Invest. 87, 383–391.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H et al. (2000). Molecular portraits of human breast tumours. Nature. 406(6797), 747–752.

Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96(16), 9212–9217.

Perou CM. (2011). Molecular stratification of triple-negative breast cancers. Oncologist 16 Suppl 1, 61-70.

Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, et al. (2006). Classification and risk stratification of invasive breast carcinomas using a realtime quantitative RT-PCR assay. Breast Cancer Res. 8(2), R23.

Petersen OW, van Deurs B. (1986). Characterization of epithelial membrane antigen expression in human mammary epithelium by ultrastructural immunoperoxidase cytochemistry. J Histochem Cytochem 34, 801-809.

Petersen OW, Polyak K. (2010). Stem cells in the human breast. Cold Spring Harb Perspect Biol 2(5), a003160.

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

Prasad CP, Rath G, Mathur S, Bhatnagar D, Parshad R, Ralhan R. (2009). Expression analysis of E-cadherin, Slug and GSK3beta in invasive ductal carcinoma of breast. BMC Cancer 9, 325.

Prat A, Perou CM. (2009). Mammary development meets cancer genomics. Nat Med 215, 842– 844.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 12, 68.

Pro B, Dang NH. (2004). CD26/dipeptidyl peptidase IV and its role in cancer. Histol Histopathol 19(4), 1345-1351.

Proia TA, Keller PJ, Gupta PB, et al. (2011). Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. Cell Stem Cell 8(2), 149–63.

Proia DA, Kuperwasser C. (2006). Reconstruction of human mammary tissues in a mouse model Nat Protoc. 1, 206-214.

Rakha EA, Reis-Filho JS, Ellis IO. (2007). Basal-like breast cancer: a critical review. J. Clin. Oncol. 26(15), 2568-81.

Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, et al. (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell 3, 109–118.

Reis-Filho JS, and Tutt AN. (2008). Triple negative tumours: a critical review. Histopathology 52, 108–118.

Reya T, Morrison SJ, Clarke MF, Weissman IL. (2001). Stem cells, cancer, and cancer stem cells. Nature 414(6859):105-11.

Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R. (2007). Identification and expansion of human colon-cancer-initiating cells. Nature 445, 111–115.

Romanska HM, Berditchevski F. (2011). Tetraspanins in human epithelial malignancies. J Pathol 223(1), 4-14.

Rønnov-Jessen L, Petersen OW, Koteliansky VE, Bissell MJ. (1995). The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. J Clin Invest 95(2), 859-873.

Rosette C, Roth RB, Oeth P, Braun A, Kammerer S, Ekblom J, Denissenko MF. (2005). Role of ICAM1 in invasion of human breast cancer cells. Carcinogenesis 26(5), 943-950.

Sadej R, Romanska H, Baldwin G, Gkirtzimanaki K, Novitskaya V, Filer AD, Krcova Z, Kusinska R, Ehrmann J, Buckley CD et al. (2009). CD151 regulates tumorigenesis by modulating the communication between tumor cells and endothelium. Mol Cancer Res 7(6), 787-798.

Seon BK, Negoro S, Barcos MP, Tebbi CK, Chervinsky D, Fukukawa T. (1984). Monoclonal antibody SN2 defining a human T cell leukemia-associated cell surface glycoprotein. J Immunol 132(4), 2089-2095.

Shackleton, M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, Visvader JE, et al. (2006). Generation of a functionalmammary gland froma single stem cell. Nature 439, 84–88.

Scharenberg CW, Harkey MA, Torok-Storb B. (2002). The ABCG2 transporter is an efficient Hoechst 33342 efflux pump and is preferentially expressed by immature human hematopoietic progenitors. Blood 99 (2), 507–512.

Sheridan C, Kishimoto H, Fuchs RK, Mehrotra S, Bhat-Nakshatri P, Turner CH, Goulet R Jr, Badve S, Nakshatri H. (2006). CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. Breast Cancer Res. 8(5), R59.

Shimano K, Satake M, Okaya A, et al. (2003). Hepatic oval cells have the side population phenotype defined by expression of ATP-binding cassette transporter ABCG2/ BCRP1. Am J Pathol 163, 3–9.

Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. (2007). Molecular definition of breast tumor heterogeneity. Cancer Cell 11(3), 259-273.

Sigurdsson V, Hilmarsdottir B, Sigmundsdottir H, Fridriksdottir AJ, et al. (2011). Endothelial induced EMT in breast epithelial cells with stem cell properties. PLoS One. 6(9), e23833.

Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD, Dirks PB. (2004). Identification of human brain tumour initiating cells. Nature 432:396–401.

Sleeman KE, Kendrick H, Ashworth A, Isacke CM, Smalley MJ. (2006). CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. Breast Cancer Res 8(1), R7.

Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, Goldhirsch A, Untch M, Mariani G, et al. (2007). 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 369(9555), 29–36.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98, 10869–10874.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 100(14), 8418–8423.

Sorrentino BP. (2002). Bcrp1 gene expression is required for normal numbers of side population stem cells in mice, and confers relative protection to mitoxantrone in hematopoietic cells in vivo . Proc Natl Acad Sci U S A 99, 12339–44.

Stendahl M, Ryden L, Nordenskjold B, Jonsson PE, Landberg G, Jirstrom K. (2006). High progesterone receptor expression correlates to the effect of adjuvant tamoxifen in premenopausal breast cancer patients. Clin Cancer Res. 12(15), 4614–4618.

Stingl J, Eaves CJ, Kuusk U, Emerman JT. (1998). Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation 63, 201–213.

Stingl J, Eaves CJ, Zandieh I, Emerman JT. (2001). Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. Breast Cancer Res Treat 67, 93–109.

Stingl J, et al. (2003). Proc Annu Meet Am Assoc Cancer Res 44: 855.

Stingl J. (2009). Detection and analysis of mammary gland stem cells. J Pathol 217(2), 229-241.

Stingl, J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, Eaves CJ. (2006). et al. Purification andunique properties ofmammary epithelial stemcells. Nature 439, 993–997.

Stipp CS. (2010). Laminin-binding integrins and their tetraspanin partners as potential antimetastatic targets. Expert Rev Mol Med 12, e3.

Stuelten CH, Mertins SD, Busch JI, Gowens M, Scudiero DA, et al. (2010). Complex display of putative tumor stem cell markers in the NCI60 tumor cell line panel. Stem Cells 28(4), 649-60.

Summer R, Kotton DN, Sun X, Ma B, Fitzsimmons K, Fine A. (2003). Side population cells and Bcrp1 expression in lung. Am J Physiol Lung Cell Mol Physiol 285, 97–104.

Taylor-Papadimitriou J, Stampfer M, Bartek J, Lewis A, Boshell M, Lane EB, Leigh IM. (1989). Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. J Cell Sci 94, 403-413.

Tomlinson GE, Chen TT, Stastny VA, Virmani AK, Spillman MA, Tonk V, et al. (1998). Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. Cancer Res 58, 3237–3242.

Turner N, Tutt A, Ashworth A. (2004). Hallmarks of 'BRCAness' in sporadic cancers. Nat. Rev. Cancer 4, 814–819.

Turner NC, Reis-Filho JS. (2006). Basal-like breast cancer and the BRCA1 phenotype. Oncogene 25, 5846–5853.

Van de Rijn M, Perou CM, Tibshirani R et al. (2002). Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am. J. Pathol. 161, 1991–1996.

Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature. 479(7372),189-93.

Vargo-Gogola T, Rosen JM. (2007). Modelling breast cancer: one size does not fit all. Nat Rev Cancer 7, 659–672.

Veneziani BM, Criniti V, Cavaliere C, Corvigno S, Nardone A, Picarelli S, Tortora G, Ciardiello F, Limite G, De Placido S. (2007). In vitro expansion of human breast cancer epithelial and mesenchymal stromal cells: optimization of a coculture model for personalized therapy approaches. Mol Cancer Ther 6(12), 3091-3100.

Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, et al. (2007). Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol 177, 87–101.

Visvader JE. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev. 23(22):2563-77.

Wang, Q, Zhang, H, Kajino, K, & Greene, MI. (1998). BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. Oncogene 17(15), 1939–48.

Watanabe K, Nishida K, Yamato M, et al. (2004). Human limbal epithelium contains side population cells expressing the ATP-binding cassette transporter ABCG2. *FEBS Lett 56, 6*–10.

Wessels LF, van Welsem T, Hart AA et al. (2002). Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. Cancer Res. 62, 7110–7117.

Wright MH, Calcagno AM, Salcido CD, Carlson MD, Ambudkar SV, Varticovski L. (2008). Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics. Breast Cancer Res 10(1):R10.

Yamazaki H, Nishida H, Iwata S, Dang NH, Morimoto C. (2009). CD90 and CD110 correlate with cancer stem cell potentials in human T-acute lymphoblastic leukemia cells. Biochem Biophys Res Commun 383(2), 172-177.

Yáñez-Mó M, Barreiro O, Gordon-Alonso M, Sala-Valdés M, Sánchez-Madrid F. (2009). Tetraspanin-enriched microdomains: a functional unit in cell plasma membranes. Trends Cell Biol 19(9), 434-446.

Yang J, Guzman RC, Popnikolov N, Bandyopadhyay GK, Christov K, Collins G, Nandi S. (1994). Phenotypic characterization of collagen gel embedded primary human breast epithelial cells in athymic nude mice. Cancer Lett 81, 117-127.

Zen Y, Fujii T, Yoshikawa S, Takamura H, Tani T, Ohta T, Nakanuma Y. (2007). Histological and culture studies with respect to ABCG2 expression support the existence of a cancer cell hierarchy in human hepatocellular carcinoma. American Journal of Pathology 170 (5), 1750–1762.

Zhang H, Somasundaram K, Peng Y, et al. (1998). BRCA1 physically associates with p53 and stimulates its transcriptional activity. Oncogene 16(13), 1713–21.

Zhang W, Li J, Allen SM, Weiskircher EA, Huang Y, George RA, Fong RG, Owen A, Hidalgo IJ. (2009). Silencing the breast cancer resistance protein expression and function in caco-2 cells using lentiviral vector-based short hairpin RNA. Drug Metab Dispos. 37(4):737-44.

Zhou, S, Schuetz, JD, Bunting, KD, Colapietro, AM, et al. (2001). The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat. Med. 7, 1028–1034.

ACKNOWLEDGEMENTS

One of the joys of the end of a journey is to look over the past and to remember all the persons who helped and supported me along this hard, long but also wonderful road.

My first debt of gratitude must go to my Tutor Prof. Francesco Salvatore to have admitted me at CEINGE and to have always trusted in me and approved some important choices that have signed my Ph.D study. I wish to express my warm and sincere thanks to my Internal Supervisor Prof. Luigi Del Vecchio to have made me discover the fascinating world of Flow Cytometry with its wide potential in both diagnosis and scientific research. I thank him for having transmitted to me his passion and wide knowledge of this scientific field and also for the appreciation and faith that he has always showed me. Finally, I thank him for having given me opportunities that have been very important to me from both a professional and a personal point of view.

I wish to thank all the members of the Flow Cytometry Group at CEINGE for their contribution in my flow cytometry training, a journey they have shared with me day by day. In particular, I thank Serena Di Cecilia for her friendship and for having shared with me an important period of my Ph.D. I also thank the members of the Cell Culture Facility Group of CEINGE to have supported me in all cell culture aspects of my research project. In particular, I thank Dr. Elisabetta Mariotti for her constant collaboration and encouragement on the first part of this thesis. Moreover, I would like to thank Prof. Bianca Maria Veneziani and Dr. Agostina Nardone - Department of Cellular and Molecular Biology and Pathology - for their collaboration on the second part of the thesis .

It is with immense and sincere gratitude that I thank my External Supervisor Dr. Stéphane Ansieau to have admitted me to his laboratory at the Cancer Research Center of Lyon, giving me the opportunity to grow and enrich myself from both a professional and

personal point of view. I thank him for his motivation, enthusiasm and immense knowledge. His guidance has helped me during a difficult and critical period of my research project. He provided the vision, encouragement and advices necessary for me to proceed through the research project and he helped me in writing of this thesis with insightful, constructive comments and questions. I thank him also for his very nice and friendly approach. I could not have imagined having a better external supervisor for my Ph.D study and I consider it an honor to have worked with him.

I thank all the members of "Failsafe programs and cellular plasticity" at the Cancer Research Center of Lyon who allowed me to be and to feel myself part of a great professional community. Their friendship and assistance has meant more to me than I could ever express. In particular, I thank Anne-Pierre Morel for her help and support on my work. I also thank Isabelle Durand, Sébastien Dussurgey and Thibault Andrieau of the Cytometry Platforms in Lyon for their support in cytometric analyses and cell sorting.

This journey would have been not the same without my dear friend Simona Langellotti. We both started and finished together our Ph.D program and we shared many personal and professional experiences that have been very important to both of us. I wish to thank her to have listened to me every time I needed and to have always been by my side with her very precious friendship. I also would like to thank her for the constant scientific interchanges and for her valuable scientific collaboration in some proteomic experiments.

I cannot find words to express my gratitude to my fiancé Jérôme Ngao whose constant love and encouragement have been essential for me during my Ph.D study. I would like to thank him also to have always been available to give me help and suggestions on my work. He already has my heart, so I will just give him a heartfelt "thanks."

Last but not the least, I wish to thank my parents, my brothers and my sister, for their love and for having always supported me in important choices throughout my Ph.D years.