Expression of Aldehyde dehydrogenase (ALDH) in the breast stem cell line D492 and relations to stem cell properties

Katrín Birna Pétursdóttir

Ritgerð til diplómaprófs
Háskóli Íslands
Læknadeild
Námsbraut í Lífeindafræði
Heilbrigðisvísindasvið

HÁSKÓLI ÍSLANDS

Tjáning á Aldehyde dehydrogenasa (ALDH) í brjóstastofnfrumulínunni D492 og tengsl við stofnfrumueiginleika

Katrín Birna Pétursdóttir

Ritgerð til diplómaprófs á meistarastigi í Lífeindafræði

Umsjónarkennari: Martha Á. Hjálmarsdóttir

Leiðbeinendur: Jón Þór Bergþórsson og Þórarinn Guðjónsson

Læknadeild

Námsbraut í Lífeindafræði

Heilbrigðisvísindasvið Háskóla Íslands

Janúar 2015

Prentun: Háskólaprent
Reykjavík, Ísland 2015

Ágrip

Aldehyde dehydrogenasar (ALDH) eru ensím tjáð bæði í eðlilegum vefja- og krabbameinsstofnfrumum. Krabbameinsstofnfrumur eru taldar viðhalda krabbameinsæxlum og stýra eiginleikum þeirra. Talið er að þessar frumur séu komnar af eðlilegum vefjastofnfrumum en öðlist illkynja eiginleika vegna uppsöfnunar á stökkbreytingum. Þær viðhalda þó einnig upprunalegu stofnfrumueiginleikunum sínum. Krabbameinsmeðferðir framkvæmdar í dag virðast ekki með nógu afgerandi hætti ráðast að krabbameinsstofnfrumum og er þá hætta á endurkomu sjúkdómsins með myndun nýs æxlis. Aukin þekking á eiginleikum krabbameinsstofnfruma er mikilvæg til að markvisst sé hægt að eyða þeim með meðferð án þess að skaða heilbrigðu frumurnar.

Í þessu verkefni voru frumulínurnar D492 og D492M skoðaðir út frá tjáningu yfirborðsprótína (markera) og ALDH ensíma í von um að geta aðgreint stofnfrumur þeirra. D492 er brjóstaþekjufrumulína með þekkta stofnfrumueiginleika. Hún myndar bæði kirtilþekju- og vöðvaþekjufrumur ásamt því að mynda þyrpingar af greinóttri kirtillíkri formgerð í þrívíðri rækt. D492M, dótturlína D492, er frumulína sem gengið hefur undir bandvefsumbreytingu þekjuvefjar (EMT) og öðlast bandvefseiginleika.

Við EMT missa frumur vel skipulagða þekjuvefs uppbyggingu sína og taka upp svipgerð bandvefsfruma með aukið frumufar. EMT á sér stað í eðlilegum fósturþroska og vefjaviðgerðum en er einnig tengt við ífarandi krabbamein, aukinn krabbameinsvöxt og meinvörp. Rannsóknir benda til þess að skörun sé á milli EMT og stofnfrumueiginleika og EMT eigi þátt í myndun og viðhaldi krabbameinsstofnfruma.

ALDH er ensím sem hvatar oxun aldehyda í karboxýl sýrur. Aldehydar geta haft skaðleg áhrif og er hvarfið þáttur í afeitrun frumunnar. ALDH virkni er talin fylgja stofnfrumueiginleikum í bæði eðlilegum og krabbameinsstofnfrumum. Megináhersla var lögð á ALDH virkni í þessu verkefni. ALDH virkni var mæld með prófi sem byggir á flúrljómandi hvarfefni sem ALDH umbreytir, myndefnið helst síðan innan í frumunni og er greint í frumuflæðisjá.

Samanburður á ALDH virkni leiddi í ljós að virknin var marktækt meiri í D492 frumum en D492M. Ræktun fruma hafði ekki marktæk áhrif á ALDH virkni skv. samanburði á mismunandi kynslóðum D492. Yfirborðsprótín sem voru prófuð sýndu yfirleitt einsleita tjáningu og voru ekki gagnleg til að spá fyrir um ADLH virkni í D492 og D492M.

D492 frumur voru flokkaðar eftir ALDH virkni og sáð í þrívíða rækt með og án brjóstaæðaþelsfruma. Frumur með háa ALDH tjáningu mynduðu þyrpingar fyrr og virtust þær hafa aukna tilhneigingu til að mynda kirtillíka formgerð en frumur með miðlungs og lága ALDH tjáningu. Þessar frumur höfðu einnig hærri tjáningu á stofnfrumumarkernum Thy-1 og fjölguðu sér hraðar en ALDH lágar frumur þegar þeim var sáð í lágum þéttleika.

Niðurstöður þessarar rannsóknar eru vísbendingar um aukna stofnfrumueiginleika D492 fruma með háa ALDH tjáningu og möguleika ALDH sem stofnfrumurmarker. Til þess að staðfesta niðurstöðurnar er þó þörf á frekari rannsóknum sem fela í sér meiri næmni á stofnfrumueiginleika.

Abstract

Aldehyde dehydrogenases (ALDH) are enzymes expressed in both normal and cancer stem cells (CSC). CSCs are hypothesized to arise from normal stem cells and acquire a malignant phenotype through series of genetic mutations. Current therapies often fail to eradicate CSCs causing metastatic relapse. Further characterization of mammary stem cells (MaSC) may increase the understanding of normal and cancer stem cells and contribute to identification of novel therapeutic targets in breast cancer.

The D492 and D492M cell lines were used in the current study. D492 is a breast epithelial cell line with stem cell properties. It generates both luminal and myoepithelial cells and forms branching colonies in 3D culture similar to structures seen in vivo. D492M is a subline of D492 that has adapted mesenchymal character through epithelial-mesenchymal transition (EMT).

EMT is a process where epithelial cells gain a mesenchymal phenotype with increased motility and lose their traditional cell-cell contacts and polar epithelial phenotype. EMT is a process occurring naturally during development and tissue repair but has also been linked to invasiveness and metastasis in cancer.

In an attempt to characterize the MaSC subpopulation, various markers were used that have been linked to normal and malignant stem cells with focus on ALDH activity.

ALDH are enzymes catalyzing the $NAD(P)^+$ dependent oxidation of aldehydes to carboxyl acids. ALDH activity is evaluated with the Aldefluor assay, where ALDH converts a fluorescence substrate to a product that is contained intracellular and detected by FACS.

In this study D492 and D492M were analyzed in terms of ALDH activity and stem cell marker expression. When compared, D492 showed higher ALDH activity than D492M. None of the selected surface stem cell marker turned out useful for the characterization and isolation of the stem cell subpopulation with high ALDH activity. A connection was suggested between ALDH activity and side scatter (SSC) that needs further analysis. Increased passage number did not significantly affect ALDH activity in D492.

D492 cells were sorted in FACSAria based on their ALDH activity into fractions containing high, medium and low ALDH levels and seeded in 3D culture (Matrigel) with and without BRENCs. Colonies formed earlier from ALDH high cells and had a stronger tendency to form branching structures. According to a western blot, ALDH high cells had increased expression of the stem cell marker Thy-1 compared to ALDH low cells and proliferated faster when seeded at a low density.

Results from this preliminary study suggest that high expression of ALDH may predict stem cell properties in D492, however further studies including additional stem cell assays are needed to verify these findings.

Acknowledgements

The project was carried out at the Department of Hematology Laboratory at Landspitali – The National University Hospital of Iceland and Stem cell research unit, BioMedical Center, University of Iceland. It was supported by the Landspitali University Hospital Research Fund.

Many people deserve acknowledgment and gratitude for their contribution to this work. I first and foremost would like to thank my supervisors Jón Þór Bergþórsson and Þórarinn Guðjónsson for creating this project and allowing me to take part in it. I am truly grateful to Jón Þór for his expert advice, intuition and technical guidance which helped form this project. I also want to thank him for being there every step of the way. Þórarinn, I would like to thank for his excellent input, providing me with materials and welcoming me to his laboratory.

I am especially grateful to Bylgja Hilmarsdóttir and Sævar Ingþórsson for all their technical assistance and guidance. Their knowledge and assistance during this time has been invaluable and truly contributed to this work.

I would also like to thank Ari Jón Arason, Eiríkur Briem and Jennifer Kricker for their assistance at the laboratory. Special thanks go to Bryndís Valdimarsdóttir for her support and answering the questions I did not dare to ask anyone else.

Sigríður Rut Franzdóttir I would to acknowledge and thank for taking the time to accompany me to Matís and assisting me with the sorting process in FACSAria. Magnús Karl Magnússon I would like to thank for his interest and advice on this project.

Last but not least I want to thank my family for their support and especially my husband Sindri Leifsson. I would not have been able to finish this project without his constant support and patience

I would like to dedicate this thesis to my grandfather Helgi Ó. Þórarinsson (1939-2014), who always encouraged me to never let anything stop me.

Table of content

Figures

Tables

Abbreviations

3D	Three dimensional
ABC	Adenosine triphosphate-binding cassette
ALDH	Aldehyde dehydrogenase
APC	Allophycocyanin
BAA	BODIPY^{TM}-aminoacate
BAAA	BODIPY^{TM}-aminoacetaldehyde
BCSC	Breast cancer stem cells
BRENCs	Breast endothelial cells
CD	Cluster of differentiation
CK	Cytokeratin
CSC	Cancer stem cells
DEAB	Diethylaminobenzaldehyde
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
ESA	Epithelial specific antigen
ESC	Embryonic stem cells
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
FSC	Forward scatter
IF	Immunofluorescence
IGF	Insulin growth factor
LEP	Luminal epithelial cells
MaSC	Mammary stem cells
MEP	Myoepithelial cells
MET	Mesenchymal-epithelial transition
MSC	Mesenchymal stem cells
MUC	Sialomucin
NOD/SCID	Non-obese diabetic/severe combined immunodeficient
PBS	Phosphate buffered saline
PE	Phycoerythrin
RA	Retinoic acid
RAR	Retinoic acid receptor
RARE	Retinoic acid response element
rBM	Reconstituted basement membrane
RPM	Revolutions per minute
RT	Room temperature
RXR	Retinoid X receptor

SEM	Standard error of mean
SSC	Side scatter
TDLU	Terminal duct lobular unit
VEGF	Vascular endothelial growth factor

1 Introduction

The molecular characterization of stem has been an ongoing project for many years with many promising results. To date no universal marker unique to all stem has been consistently reported. The research focus in this field seems to have shifted from elucidating common stem cell surface markers to metabolic markers, relevant to maintaining the stem cell state and/or regulating proliferation and differentiation (4). An example of such a marker is aldehyde dehydrogenase (ALDH) activity. Normal and malignant stem cells have been consistently shown to express high ALDH levels (5). In the current study the breast epithelial cell line D492 was used to explore the potential of aldehyde dehydrogenase as a marker of stem cell properties.

1.1 Human mammary gland

The mammary gland is a dynamic organ that can be viewed as; a modified sweat gland comprised of epithelium and surrounded by stroma. The gland consists of series of branching ducts that terminate in the terminal duct lobular unit (TDLU). The ductal tree is composed of two types of epithelial cells; an inner layer of luminal epithelial cells (LEP) and an outer layer of myoepithelial cells (MEP) that reside on a basement membrane (Figure 1). TDLU is the functional unit of the gland composed of lobules with acini where the luminal epithelial cells produce milk which is transported to the nipple by contraction of myoepithelial cells (6-8).

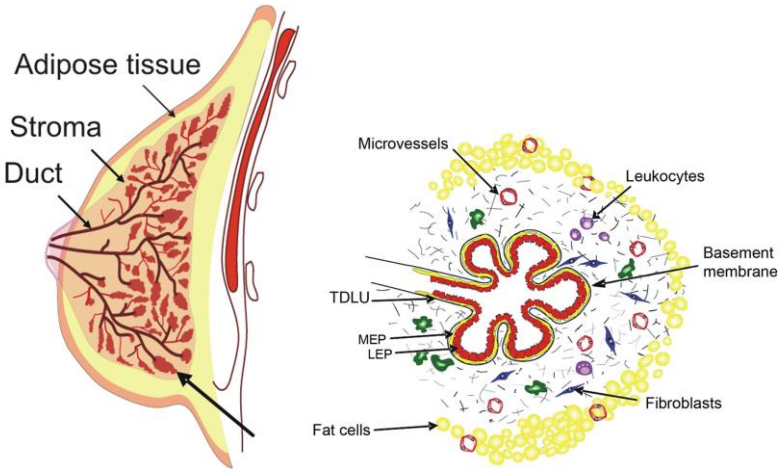


Figure 1 Schematic picture of the human breast and TDLU

The gland is built up of ducts that terminate in TDLU composed of luminal epithelial and myoepithelial cells and surrounded by a stroma. Adapted from Sigurdsson, 2005 (1)

1.2 Stem cells

Stem cells are undifferentiated cells that give rise to all other cells in the human body at all stages; embryonic, fetal and adult (9).

1.2.1 Origin of stem cells – a short overview

The term “stem cells” originally appeared in scientific literature in 1868 in book chapters written by the German scientist Ernst Haeckel (10, 11). It is a direct translation of the German word “Stammzellen”. Haeckel stated the term to be simple and appropriate as it entails that all cells are derived from the stem cell to form a multicellular organism. He claimed a fertilized egg cell to be the original stem cell with partial origin from both father and mother (10).

Theodore Boveri, a student of Haeckel later reported that stem cells from a fertilized egg gave rise to primordial germ cells, and from them branched off primordial somatic cells. He introduced the characteristics of stem cells; self-renewal and differentiation to somatic and germ cells (10).

The existence of hematopoietic stem cells was first recognized by Till and McCulloch almost a century after Haeckel introduced the term. In their experiment they found colony forming cells, which were clones originally and most likely derived from one cell (12). That one cell was thus capable of forming multi-lineage hematopoietic colonies (12, 13). The stem cells were considered to be found in most normal tissue but presumably maintained as a rare population (14).

1.2.2 Stem cell categorization

Stem cells are categorized as totipotent, pluripotent and multipotent.

Totipotent stem cells can be described as the cells on top of the pyramid. They represent the most undifferentiated stem cells. Totipotent stem cells are present in early development and form both the embryo and the placenta (9).

Pluripotent stem cells are able to differentiate to cells of all germ layers; ectoderm, mesoderm and endoderm which all organs and tissues are generated from, they form the embryonic stem cells (ESC) (9, 15).

Multipotent stem cells are located in most tissues. Their differentiation capacity is more limited than pluripotent cells as they can naturally only differentiate to cells within a distinct germ layer. They give rise to various cell lineages. Multipotent cell can be further categorized to oligopotent, bipotent and unipotent depending on the number of cell lineages they can give rise to. Bipotent for example which as the name entails can only generate two cell lineages for example luminal epithelial and myoepithelial cells in the mammary gland (9).

1.2.3 Stem cell properties and characteristics

The stem cell is foremost defined by properties like self-renewal, the ability to differentiate to various cell types (potency) and single cell origin (clonality). These characteristics can however apply differently to various stem cell types (9).

Only stem cells self-renew by a highly regulated symmetric or asymmetric division (16). During symmetric division the stem cell produces two cells identical to the original stem cell or two committed progenitor cells that differentiate. Such a division expands the stem cell population or decreases it, respectively. By asymmetric division the stem cell produces one stem cell and one progenitor cell leaving the stem cell population unchanged. How the balance between symmetric and asymmetric

division is regulated is still poorly understood (17). The self-renewal ability varies between cell types. Multipotent stem cells from blastocyst show a great ability to self-renew while stem cells in adult tissue do not (9).

Differentiation capacity depends on the stem cell type and its origin. ESCs are able to differentiate to all possible cell types. In contrast cells categorized as tissue-specific stem cells, like mammary stem cells (MaSC), are only able to differentiate to cell types in a given tissue and therefore restricted to a particular organ (16).

1.2.4 Mammary stem cells (MaSC)

The existence of mammary stem cells (MaSC) is underlined by dynamic cellular changes occurring, at puberty and during pregnancy where the gland fully develops just prior to lactation. These changes involve active proliferation and differentiation of epithelial cells when the TDLU form (puberty) and the gland increases in to at least 4-fold the normal size (pregnancy) (6).

There is evidence for hierarchy with respect to cells of the breast epithelium ranging from stem cells to differentiated cells; however this cell hierarchy is poorly understood. The MaSC are defined by their ability self-renew and to generate both ductal and lobular components of the mammary epithelium with all the cell types it entails (18).

The MaSC are proposed to be of luminal epithelial lineage. They are so called supra-basal, with no luminal contact but are distinct from myoepithelial cells. The cells are positive for the luminal marker; epithelial cell adhesion molecule (EpCAM, also known as epithelial specific antigen – ESA) but negative or weakly positive for sialomucin (MUC1), a marker of fully differentiated luminal epithelial cells. These cells have been shown to give rise to both luminal and myoepithelial cells. Myoepithelial cells can however not generate luminal epithelial cells (7, 19). MaSC reside in the mammary ducts rather than the lobules as demonstrated by stem cell marker expression and mammosphere forming ability (20).

Elucidating MaSC existence in mice has been relatively successful as compared to experiments using human models. Shackelton et al. confirmed the existence of MaSC in mice by generating a fully functional mammary gland from a single stem cell expressing the surface markers; $Lineage^-$ (Lin, markers of normal cell types: CD2, CD3, CD10, CD16, CD18, CD31, CD64 and CD140b), $CD29^{high}$ and $CD24^+$ (21).

1.2.5 Cancer stem cells (CSC)

The hypothesis of cancer stem cells (CSCs) and their contribution to the origin of cancer has been around for quite some with numerous speculations and extensive debate. The first report of the hypothesis is dated back to mid-18^{th} century when R. Virchow, a German pathologist, predicted that cancer originated from embryonic-like cells in developed tissue (22, 23).

The hypothesis states that tumors arise from a subpopulation of stem cells (or early progenitor cells) termed as tumor initiating cells or cancer stem cells. For cancer to occur, a series of genetic mutations must take place within a cell. Stem cells have the capacity to generate progeny with many

cell divisions and have a long lifetime. This makes them likely to accumulate genetic mutations over a long period of time, due to persisting genetic instability or environmental factors (17, 18, 24). A likely consequence of genetic mutations is the dysregulation of the tightly regulated self-renewal process of tissue stem cells leading to malignant transformation to CSCs (18, 25). Another possibility is that CSCs originate in more differentiated cells such as transit amplifying cells. Genetic and/or heterotypic changes in these cells cause them to acquire stem-like properties and to function as CSCs (26). Most likely both events can occur depending on the cancer type.

The American Association of Cancer Research (AACR) established a much needed definition for CSCs. It states that a CSC "is a cell within a tumor possessing the capacity to self-renew and giving rise to heterogeneous lineages of cancer cells comprising the tumor". Accordingly, the cells are defined experimentally by their ability to form a new tumor with the same phenotypic heterogeneity as the original tumor (27). CSCs and normal stem cells share important properties which can be useful in characterization and isolation of CSCs. These properties include; self-renewal, differentiation ability, active telomerase expression, activation of anti-apoptotic pathways, increased membrane transporter activity, migration and metastasizing ability and anchorage independence (mammosphere forming ability in breasts) (16, 17).

Even though the hypothesis was first stated in the mid-18th century it took scientists well over a century to support it by experimental data and more evidence is accumulating as time passes and assays improve. One of the first experimental evidence showed that human leukemia is driven by leukemic stem cells. Isolated and transferred leukemic stem cells were able to recapitulate the disease phenotype in non-obese diabetic/severe combined immunodeficient (NOD/SCID) mice (28, 29). The leukemic stem cells expressed the same stem cell markers ($CD34^+$ and $CD38^-$) as normal non-malignant stem cells supporting the hypothesis that the malignancy originated from a normal stem cell (29).

Evidence of cancer stem cells in the human breast tumors was demonstrated by Al-Hajj et al. They isolated $CD44^+CD24^{lo/-}$ cells from human breast tumors which were able to generate a tumor in NOD/SCID mice with the same heterogeneity as the initial tumor (24). In addition, cancer stem cells have been identified in cancer of the brain (30), colon, pancreas, lung, prostate and melanoma (26). In the attempt to characterize CSC in tumors various markers have been used including CD24, CD44, CD29, CD133, EpCAM (ESA), Thy-1 and ALDH activity (24, 26, 28, 30). The markers are also known to be expressed in normal tissue. These markers were used in this study for the characterization of the stem cell population in D492. A marker review is found in the appendix.

CSCs have been shown to be relatively resistant to chemotherapy and radiation. It has been suggested that this resistance is due to their proliferation rate, which causes cell-cycle active chemotherapeutic agents to be less effective, as they primary target highly proliferative transit amplifying cells that make up the bulk of the tumor. CSCs may also possess a set of highly active adenosine triphosphate-binding cassette (ABC) proteins that cause an efflux of chemotherapeutic agents. Another reason can be attributed to metabolization of chemotherapeutic drugs e.g. through high expression of ALDH enzymes (25). Dramatic changes in the gene expression pattern of tumor cells after treatment suggest that additional factors are likely to impact CSC resistance (31). Studies

have shown that multiple signaling pathways play a role in drug resistance e.g. the EGFR/PI3K/PTEN/Akt/mTORC pathway which is altered due to mutations or abnormal gene expression (32). Overall the resistance to chemotherapeutic drugs can be attributed to the ability of CSCs to survive the treatment and implement a relapse by initiating a new tumor. In view of this, there is an increasing demand for therapeutics that target both CSC and tumor cells.

There are indications that CSCs have the ability to carry out conserved developmental programs including Epithelial-mesenchymal transition (EMT). At least, the EMT signaling networks appear to be active during CSC development and maintenance (33).

1.3 Epithelial-mesenchymal transition (EMT)

1.3.1 Normal EMT

Epithelial-mesenchymal transition (EMT) is a process occurring normally during development. Polarized epithelial cells are reprogrammed to mesenchymal cells displaying a motile phenotype. Epithelial cells are closely attached and arranged via desmosomes, adherens- and tight junctions. Mesenchymal cells are albeit much loosely organized and form a three-dimensional extracellular matrix. In the EMT process, the contact between epithelial cells is loosened and cell polarity is lost, enabling cell motility and acquisition of invasive properties. EMT along with the reverse process, Mesochymal-epithelial transition (MET), is essential for gastrulation and organogenesis (34, 35).

In normal adult tissue, EMT is in a dormant state but activated for tissue regeneration and wound repair (34).

1.3.2 Abnormal EMT

When EMT is deregulated it can cause pathological conditions as in organ fibrosis. Moreover EMT has been highly linked to cancer progression, increased invasiveness and metastasis (35). During metastasis, epithelial tumor cells acquire a mesenchymal phenotype through EMT. The highly motile EMT cells evade the basement membrane and surrounding tissue to enter the blood stream or lymphatic system. In order to inhabit a new location the cancer cells must undergo MET to form macrometastasis, a secondary tumor (35, 36). Therefore EMT correlates with cancer relapse, poor survival rate and clinical outcome (35).

EMT can be induced by multiple signaling pathways, regulators and effector molecules (36). These pathways include Wnt, TGF-β, Hegdehog and Notch. Eventually these pathways influence key transcription factors involved in EMT induction like Snail, Twist, Slug and ZEB1 (33, 36). As cells go through EMT, expression of certain markers can change through among other ways; transcriptional control where epithelial marker expression is reduced or lost (E-Cadherin and cytokeratins) and expression of mesenchymal markers is increased (e.g. N-Cadherin, Vimentin and fibronectin) (37).

EMT has been proposed to be a process which increases stemness as cells that have undergone EMT aquire a $CD44^+$ and $CD24^-$ phenotype and have increased ability to form mammospheres. Mani et al. demonstrated this by inducing EMT in both normal and malignant breast epithelial cells. EMT in

normal cells generated cells with stem like properties that formed mammospheres and cancer cells gained stem like properties and were able to initiate tumors (38).

ALDH activity also a marker for stemness has however been shown not to be increased in cells that have undergone EMT (39, 40). Sarrio et al. demonstrated that spontaneous EMT in normal basal breast epithelial cells does not increase all stem cell properties. They hypothesize that EMT in both normal and cancer breast cells represents incomplete myoepithelial differentiation, rather than a shift towards a more stem cell like state. Accordingly, instead of viewing EMT as a stem cell generating program it can be seen as a process increasing phenotypic and functional diversity in normal and malignant breast epithelium (39).

Further Liu et al. studied the ability of breast cancer stem cells (BCSC) to shift between EMT and MET states. They demonstrated that BCSC that have undergone EMT show a gene-expression profile more similar to basal stem cells in a normal mammary gland. These are $CD44^+$ and $CD24^-$, mostly quiescent and located at the invasive tumor front close to the stroma. In contrast, BCSC that have undergone MET show gene-expression profiles like luminal stem cells. These cells express ALDH, are more proliferative and inhabit tumors center rather than the edges (40).

A relationship between EMT and metastatic potential has been recurrently suggested. According to some studies, cells expressing $CD24^-$ and $CD44^+$ attend to possess more invasive capacity than ALDH expressing cells however; both types of cells are more invasive than cells from the bulk of the tumor. The strongest invasive potential is obtained by selecting for both $CD24^-/CD44^+$ and ALDH high cells (40).

1.4 Aldehyde dehydrogenase (ALDH)

Aldehyde dehydrogenasse are enzymes with an extensive and complex function. There are currently 19 known isoforms (isozyme) encoded by 19 different genes in humans. The enzymes have been located in the nucleus, cytoplasm, endoplasmic reticulum and mitochondria, with some of them found in more than one location. The enzymes show a wide tissue distribution and are substrate-specific (41, 42). Mutations in the human ALDH genes are linked to various diseases including Sjögren-Larsson syndrome, type II hyperprolinemia, γ-hydrobutyric aciduria and cancer (43).

1.4.1 General ALDH function

The general function of ALDH is to catalyze the $NAD(P)^+$ dependent irreversible oxidation of aldehydes to carboxyl acids. Aldehydes are a product of various metabolic processes from both endogenous and exogenous precursors. Over 200 different types of aldehydes may result from oxidative degradation of membrane lipids (lipid peroxidation) (42). Aldehydes are also generated from metabolism of amino acids, biogenic amines, carbohydrates, vitamins, steroids and neurotransmitters (3, 43). In addition, aldehydes are formed through biotransformation of numerous drugs and environmental agents (43).

Most aldehydes have cytotoxic effects. They form adducts with different cellular targets and the consequences can be; interrupted cellular homeostasis, inactivation of enzymes, DNA damage and

cell death. Therefore, conversion of aldehydes by ALDH to carboxyl acids in many cases serves to detoxify harmful substances and protects the cell (41, 42).

However, the generation of carboxyl acids, may also be essential for normal cellular function. Example of that is retinoic acid (RA), formed by oxidation of retinal (retinaldehyde) originally from retinol (vitamin A). RA plays a role in normal growth, differentiation, apoptosis and cell cycle arrest in addition to development and maintenance of epithelial tissue. The function is carried out by regulating gene expression. RA diffuses into the nucleus where it serves as ligand for RA receptor (RAR) and retinoid X receptor (RXR). The receptor complex (RAR and RXR) then binds to retinoic acid response elements (RAREs), within target gene regulatory sequences thereby inducing transcription (Figure 2) (3, 42).

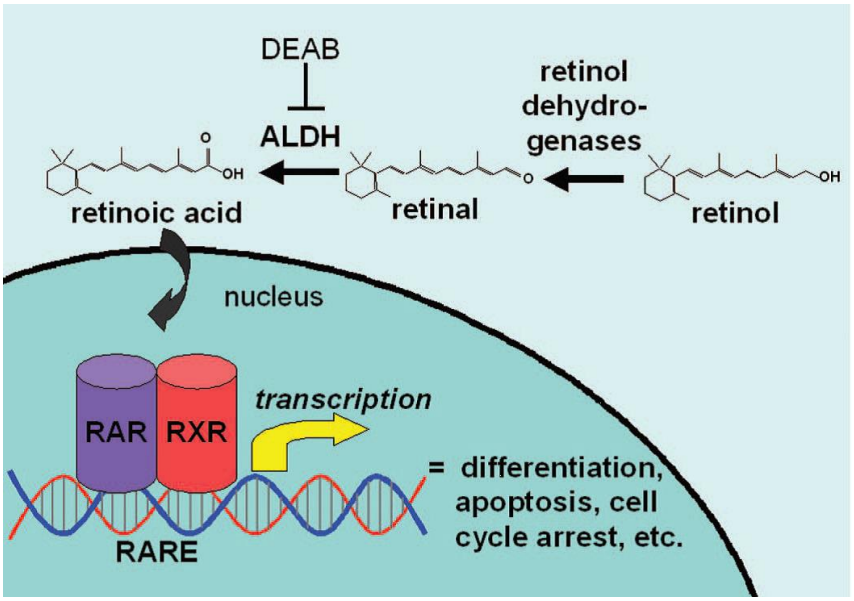


Figure 2 Function of ALDH in retinoic acid (RA) production and signaling

ALDH converts retinal to RA which diffused through the nucleus and works as a ligand for RAR and RXR which in turn bind to RARE with transcriptional effects. Adapted by Marcato et al, 2011 (3).

1.4.2 The ALDH superfamily

The genes coding for ALDH belong to one superfamily. They are well conserved and considered to derive from a common gene ancestor (homologous) dating back to 3 billion years ago. The superfamily is categorized to 24 families in eukaryotes. Only 11 families are found in humans, 1-9, 16 and 18 which together contain the 19 ALDH genes located on various chromosomes. Each family, which includes up to six genes, is further categorized to subfamilies (41).

The classification of genes to families and naming is based on a nomenclature system published in 1999. Proteins derived from genes of the same family share over 40% of the amino acid sequence. Proteins from genes within the same subfamily share over 60% (41).

1.4.3 ALDH1A subfamily

Members of the subfamily ALDH1A (ALDH1A1, ALDH1A2 and ALDH1A3) are quite closely related, with ALDH1A1 and ALDH1A3 known to share 70% of the amino acid sequence (44). The enzymes are all located in the cytosol of the cell. They all share the same main function; to catalyze oxidation of all-trans-retinal and 9-cis-retinal to RA and thus regulate RA signaling (41, 42). These enzymes are sometimes referred to as RALDH (retinal dehydrogenase) (45).

1.4.3.1 ALDH1A1

ALDH1A1 is a homo-tetramer located in adult epithelium cells, in several organs including brain, liver, kidney, lung, eye lens, retina and testis (42). ALDH1A1 is also expressed in hematopoietic stem cells (HSC) and influences differentiation by producing RA. For therapeutic purposes, it is possible to promote HSC expansion by inhibiting ALDH1A1 (46).

The participation of the enzymes in oxidation of retinal to RA is clearly demonstrated when the *ALDH1A1* gene is disrupted in mice. The mice are viable but show higher levels of serum retinal with reduced RA synthesis in the liver (42, 47).

ALDH1A1 plays a role in defending the cell against oxidative stress by oxidizing LPO-derived aldehydes but also by reducing the oxidation of free fatty acids and limiting the production of reactive oxygen species (ROS) (42).

The enzyme seems to be important when it comes to cancer therapeutics. It is capable of protecting the cell from anti-cancer drugs like cyclophosphamide (CP) by detoxifying the aldehydes. The cell becomes insensitive to the drug so its effectiveness is highly reduced (48). This effect may potentially explain the drug resistance possessed by CSCs.

High ALDH1A1 expression mediated by CSCs was demonstrated to be a predictor of metastasis and poor clinical outcome in breast cancer (49). However others failed to support this result and suggested another isoform to be responsible like ALDH1A3 (44).

1.4.3.2 ALDH1A3

ALDH1A3 is a homo-dimer expressed in breast, stomach, kidney, salivary glands and fetal nasal mucosa with a functional role in embryonic development (42). Unlike *ALDH1A1,* disruption of *ALDH1A3* function is lethal. Knockout mouse fetuses die at birth due to severe malformations of the nasal cavities which causes lethal respiratory distress syndrome (50).

Like ALDH1A1, ALDH1A3 influences oxidative stress and minimizes cellular damage by oxidizing and therefore detoxifying LPO-derived aldehydes (42).

ALDH1A3 deficiency has been linked to various cancer types for instance breast cancer where a causal influence on RA synthesis is suspected (51). However Marcato et al. correlated expression of ALDH1A3 with high tumor grade in addition to metastatic in breast cancer. ALDH1A3 expression was shown to be superior to ALDH1A1 expression as a predictor of unfavorable clinical outcome (28, 44).

1.4.4 ALDH activity as a stem cell and CSC marker

Cells expressing high levels of aldehyde dehydrogenases have been identified in populations of hematopoietic, neural, mesenchymal, endothelial and mammary stem and progenitor cells. Supporting the idea that ALDH activity might be a good marker of stem cell properties in different types of normal and malignant tissue (5, 52).

Cheung et al. were among the first to display increased ALDH activity ($ALDH^+$) in leukemic stem cells derived from bone marrow samples of acute myeloid leukemia (AML) patients (14 out of 43 samples). The $ALDH^+$ cells co-expressed CD34 (a leukemia stem cell marker) and displayed a much more robust leukemic engraftment in NOD/SCID mice than $ALDH^-$ cells (3, 53).

An analysis by Ginester et al. further elucidated the function of ALDH as a marker for stem and progenitor cells in both normal and malignant breast. They revealed that $ALDH^+$ cells isolated from normal breast tissue was able to generate mammospheres while $ALDH^-$ cells did not. $ALDH^+$ cells from breast cancers where able to generate tumors in humanized cleared fat-pads of NOD/SCID mice. A tumor was formed even with only 500 cells seeded. Moreover, the tumors displayed the same phenotypic heterogeneity as the original tumor, with a similar $ALDH^+/ALDH^-$ ratio. However ALDH- cells did not possess tumor formation ability regardless of the amount of cells injected. $ALDH^+$ with the additional expression of $Lin^-/CD44^+CD24^-$ was displayed as a very small population but had high tumorigenic capacity in assays and formed a tumor with only 20 cells injected. In contrast $ALDH^-$ cells with the $Lin^-/CD44^+CD24^-$ phenotype were not tumorigenic (28).

These findings are the fundamental for the establishment of ALDH as a stem cell marker in normal and malignant tissue. However ALDH as a marker of stem cells is still controversial. ALDH activity was demonstrated by Eirew et al. as a marker for luminal progenitor cells instead of stem cells. MaSC and primitive progenitor cells displayed a low ALDH activity that became elevated as the cells committed to the luminal lineage (54).

1.5 Aldefluor assay

ALDH activity of cells can be measured with the Aldefluor assay. As substrate, the assay utilizes BODIPY aminoacetaldehyde (BAAA). BODIPY is a fluorochorme and bound to aminoacetaldehyde it forms a fluorescent substrate that diffuses freely into cells without being toxic (commercial name Aldefluor). BAAA is converted to a carboxylate ion, BODIPY aminoacate (BAA), in the presence of ALDH and contained intracellular. The fluorescence measured is proportional to the amount of active ALDH found in the cell. ALDH is inhibited by diethylaminobenzaldehyde (DEAB) which serves as a negative control, to confirm that fluorescence is due to ALDH activity. The obvious advantages of this method are its simplicity, safety and minimal toxic effects on cells. The fluorochrome BODIPY is detected by visual light, does not require UV excitation and has no significant overlap with other fluorochromes. The assay can therefore be combined with detection of other markers (55).

The assay was originally developed and optimized for cells of hematopoietic origin including hematopoietic stem cells (HSC) (55). However, multiple reports have underlined its relevance in other cell types, both normal and cancerous, e.g. mammary epithelial cells (28).

The HSCs express high levels of ALDH1A1. Some studies have suggested that ALDH1A1 activity is mostly responsible for positivity in the Aldefluor assay (3, 42, 55). In contrast, later studies did not confirm this and concluded that ALDH1A1 alone was not responsible for the activity detected. Measurements of HSC with ALDH1A1 deficiency showed no reduction in ALDH activity detection. This suggests that other isoforms may be contributing to the ALDH activity detected, possibly ALDH2 and ALDH3A1 which were notably expressed (56). In addition it seems that the ALDH activity detected is not primarily due to ALDH1A1 in breast cancer tumors but rather ALDH1A3 among other isoforms (3).

ALDH isoforms have been shown to display substrate-specificity but can nonetheless cross-react, meaning that more than one isoforms can process the same substrate. It is difficult to map which of the isoforms contribute to the ALDH activity detected, it most likely varies and depends on the cells origin and cancer type (3).

1.6 D492 and D492M

D492 is a breast epithelial cell line with stem cell properties created by Gudjonsson et.al (7). The cells were isolated and immortalized from primary cultures of biopsies from reduction mammoplasty, meaning that they should represent healthy normal breast tissue. D492 is a multipotent cell line with basal-like phenotype that expresses; EpCAM (ESA), claudin-1 and cytokeratins CK8, CK18 and CK19 which are luminal markers. It also expresses the myoepithelial markers CK5/6 and CK14 (7, 37).

The stem cell properties of the cell line are evidenced by their ability to differentiate into both luminal and myoepithelial cells. The cells are also capable of forming branching TDLU-like structures when embedded in a three-dimensional (3D) reconstituted basement membrane (rBM). It has been proposed that EpCAM (Epithelial cell adhesion molecule) expressing cells, but not MUC (sialomucin - a luminal marker) expressing ($EpCAM^+/MUC^-$) are the precursor cells of TDLU and breast stem cell candidates. The cells are proposed to be situated among the myoepithelial cells on the basement membrane but with no luminal contact despite their luminal lineage origin (Figure 3) (7, 37).

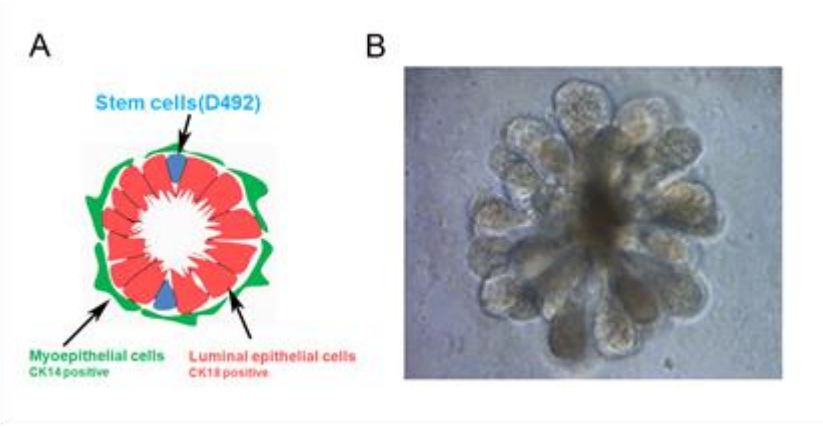


Figure 3 D492 a breast epithelial cell line with stem cell properties

A: The stem cells of D492 are likely to be situated on the basement membrane among the myoepithelial cells with no luminal contact (schematic picture of a TDLU). **B:** In 3D culture D492 form TDLU like structured colonies**.** Adapted from Sigurdsson, 2005 (1).

D492 forms a monolayer in culture and displays high trans-epithelial resistance when cultured on transwell filters (7).

D492M (mesenchymal) is a subline created by Sigurdsson et.al (37). It is derived from D492 by the isolation of spindle-like colonies grown in co-culture with breast endothelial cells (BRENCs). The cells adapted an EMT phenotype and express high levels of Vimentin, N-Cadherin and alpha-smooth muscle actin, while expression of E-Cadherin and cytokeratins 5/6, 8, 14, 17 and 19 is lower compared to D492. Thy-1 (CD90) has been shown to be highly up-regulated in D492M (37).

1.7 Three dimensional (3D) culture

1.7.1 Reconstituted basement membrane (rBM) - Matrigel

Presumably 3D cell culture models have the ability to stimulate natural conditions in more detail than possible with conventional monolayer cultures (Figure 4). Under such conditions mammary epithelial cells form ducts, ductules and glandular-like structures with a lumen to which cells can secrete casein (57).

Typically, reconstituted basement membrane (rBM) is used for 3D culture known by its commercial name as Matrigel. Matrigel is made from extracts of a tumor, named after Engelbreth-Holm and Swarm (EHS). The tumor was discovered in mouse and is most likely derived from parietal endoderm. It has abundance of extracellular matrix that resembles a basement membrane (58).

D492 has previously been shown to form TDLU structures when embedded in Matrigel

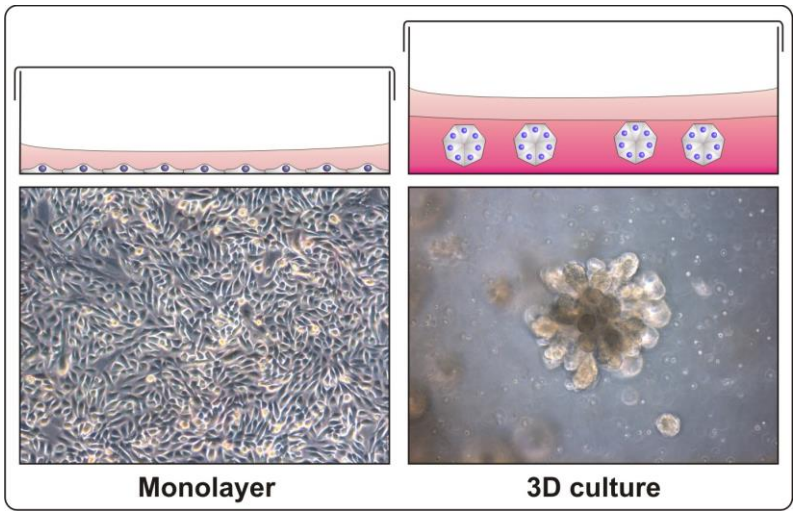


Figure 4 A schematic and representative pictures of cells in monolayer and 3D culture.

The difference between cells grown in monolayer and 3D culture. Adapted from Ingthorsson, 2014 (2).

1.7.2 Mammosphere assay

The mammosphere assay, also referred to as non-adherent culture, was originally modified from a neurosphere assay. The mammosphere assay is used to quantify stem and early progenitor cells activity and to demonstrate their self-renewal ability (59).The mammosphere assay is based on the idea that stem cells (or early progenitor cells) are viable without cell contacts while other types of cells die due to anoikis (60). The viable cells self-renew and proliferate to form a sphere. The assay is carried out by seeding cells in ultra-low attachment plates in medium and stemness is then determined on the basis of spheres formed over a predefined time interval (59).

2 Aim of the study

The aim of the study is to characterize the stem cell population in D492, a breast epithelial cell line, and compare it to its mesenchymal-like subline D492M. The characterization will entail various markers (including CD24, CD44, CD90, CD117, CD133) already demonstrated to be useful in stem cell isolation. However, intense focus will be on a relatively recently discovered metabolic marker Aldehyde dehydrogenase (ALDH). It will be determined whether or not high ALDH levels reflect stem cell phenotype such as ability to form colonies in 3D culture and/or non-adherent culture. In addition, the purpose is to see if ALDH levels remain constant over time by comparing different passages of D492.

This study has the potential to increase understanding of normal and cancer stem cells and may contribute to identification of novel therapeutic targets. The eventual goal is eradication of CSCs without discernably affecting the normal stem cells.

3 Material and methods

3.1 Culture

The cell lines D492, D492M and breast endothelial cells (BRENCs) were all cultured in T25 culture flasks in a monolayer. D492 and D492M in collagen-1 (Inamed Biomaterials) coated flasks. All mediums used for culture was supplemented with 50 IU/ml of penicillin and 50 µg/ml of streptomycin (Sigma).

D492 and D492M were maintained on H14 medium. H14 consists of DMEM:F12 basal medium (Gibco) supplemented with growth factors; insulin (final conc. 250 ng/ml), transferrin (final conc. 10 µg/ml), EGF (final conc. 10 ng/ml), Na-Selenite (final conc. 2,6 ng/ml), Estradiol (final conc. 10^{-10}M), Hydrocortisone (final conc. $1,4*10^{-6}$M) and Prolactine (final conc. 0,15UI). The growth factors were purchased from Sigma, except EGF from Peprotech and Na-Se from BD Biosciences.

BRENCs were maintained on EGM30 medium with 30% fetal bovine serum (FBS). The EGM medium is comprised of EBM-2 basal medium (Lonza) supplemented with growth factors of unknown concentration; hydrocortisone, bFGF, EGF, VEGF, IGFR3, Heparin and absorbic acid (Lonza). EGM5 medium contains the same growth factors but only 5% FBS.

All cell lines were cultured in an incubator set at 37°C with 5% CO_2.

3.2 Cell preparation for immunophenotyping with FACS

The cells were trypsinized with 1 ml 0.25% trypsin/EDTA for around 3-5 minutes. Subsequently the trypsin was inactivated by adding approximately 3 ml of 10% FBS. The cells were centrifuged for 4 minutes at 2000 RPM, supernatant discarded and cells re-suspended in PBS.

3.3 Immunophenotyping with FACS

Cells were stained with antibodies conjugated with different fluorochromes (Table 1). For each antibody 3-5 µL were used for cell density of 1 million cells/ml or less. Cells mixed with antibodies were incubated for 30 minutes at room temperature (RT) in the dark. Cells were washed with 1 ml of PBS and centrifuged for 4 minutes at 2000 RPM to wash away unbound antibodies. Cells were re-suspended in PBS or fixed with 0.4 ml of Cellfix (BD Bioscience).

Table 1 Information on antibodies for FACS

Antibody	Clone	Species	Isotype	Conjugated fluorochrome	Company Order number
CD24	ML5	Mouse	IgG2a, κ	PE	BD Bioscience 560991
CD29	TS2/16	Mouse	IgG1κ	APC	Miltenyi Biotec 130-101-271
CD44	G44-26 (C26)	Mouse	IgG2b, κ	PerCP-Cy5.5	BD Bioscience 560531
CD90 (Thy-1)	DG2	Mouse	IgG1	APC-vio770	Miltenyi Biotec 130-099-289
CD117	104D2	Mouse	IgG1	PE	BD Bioscience 332785
CD133/2	293C3	Mouse	IgG2b	PE-vio770	Miltenyi Biotec 130-104-155
CD324 (E-Cad)	674A	Mouse	IgG1	PE-vio770	Miltenyi Biotec 130-101-095

3.4 Aldefluor assay

The Aldefluor assay kit from STEMCELL Technologies (Vancouver, British Columbia, Canada) was used to evaluate ALDH expression. Cells were prepared for the assay in the same way as for immunophenotyping. After centrifugation, supernatant was removed and cells re-suspended in 1 ml of ALDEFLUOR™ Assay Buffer. Next, 5 µl of ALDEFLUOR™ Reagent was added and mixed with 1 ml cell suspension (never greater than 1 million cells/ml). Immediately, 0.5 ml of this mixture was transferred to a tube containing 5 µl of ALDEFLUOR™ DEAB Reagent (negative control). Both sample and control were incubated at 37°C for 30-45 minutes and centrifuged for 4 minutes at 2000 RPM. Following centrifugation, supernatant was discarded and cells re-suspended in 0.5 ml of ALDEFLUOR™ Assay Buffer. Cells were kept chilled on ice prior to analysis.

When additional immunophenotyping was performed, cells were stained with antibodies for 15-30 minutes in the dark and kept on ice. After staining they were centrifuged in a cooled centrifuge (4°C) for 4 minutes at 2000 RPM. The cells were then suspended in 0.5 ml of ALDEFLUOR™ Assay Buffer and analyzed with FACS as soon as possible. The cells were kept on ice in order to slow down the product (BAA) efflux. Importantly, the ALDEFLUOR™ Assay Buffer was used at all times during the assay as it contains ABC transporters activity inhibitors to minimize efflux of both substrate and reaction product (61).

3.5 Flow cytometry / FACS (Fluorescence-activated cell sorting)

Flow cytometry or FACS was performed using the instrument MACSQuant from Miltenyi Biotec (Bergisch Gladbach, Germany). The instrument was compensated for each fluorochrome and settings

adjusted for the analysis. Calibration was performed with MACSQuant calibration beads (Miltenyi Biotec).

Analysis of FACS data was carried out by using the program Flowjo, version 8.8.6. To analyze data a live gate was created around the cell population using the Forward Scatter (FSC) and Side Scatter (SSC) parameters, in order to gate out cell debris, clustered cells and artifacts. ALDH fluorescence was analyzed with ALDH/SSC dot plots, including live gate events. In this plot a gate was set at the right edge of the stained population using negative controls. This gate was subsequently used for the definition of ALDH activity in the corresponding samples.

3.6 Cell sorting

Cell sorting was performed on FACSAria™ II from BD (Franklin Lakes, NJ, USA) located at Matís ltd. (Icelandic Food and Biotech R&D). Prior to sorting, cells were prepared for staining as done before with the additional step of filtering the cells through a 30 µm filter after centrifugation and re-suspension. The cell suspension was adjusted to 2 million cells/ml, and stained with 5 µl of ALDEFLUOR™ Reagent per ml of cell suspension. The assay was carried out as previously. After staining the cell suspension was adjusted to 5 million cells/ml.

Sorting gates were set for 10% of ALDH high cells (P1), 48% ALDH medium cells (P2) and 10% ALDH low cells (P3). ALDH high and low cells were sorted simultaneously and the ALDH medium cells thereafter. The sorted cells were collected in 4 ml glasses containing 1 ml of EGM5 medium.

3.7 Protein isolation

Protein was isolated from ALDH sorted high, medium and low cells. The cells were originally seeded on an untreated polystyrene 12 well plate (Falcon), 2000 cells per well and cultured for 10 days on EGM5. Cells in each well were washed with PBS and treated with 60 µl of RIPA lysis buffer (contains for each ml of RIPA: 10 µl protease inhibitor, 10 µl phosphatase inhibitor and 10 µl EDTA) for 10 minutes. The bottom of the wells was scraped with a specific cell scraper and cells transferred to an eppendorf tube. Tubes were incubated for 10 minutes, frozen in liquid nitrogen and subjected to five freeze-thaw cycles. After centrifugation for 20 minutes at 4°C and at 12.000 RPM, supernatant was transferred to a new eppendorf tube and frozen at 20°C. All steps of the protein isolation were carried out on ice.

3.8 Western blot

For Western blot 5 µl protein sample was mixed with 2.5 µl of NuPAGE sample buffer (Novex by Life technologies), 1µl mercaptoethanol and deionized water up to 10 µl. The mixture was heated on a heat plate at 98°C for 10 minutes. Samples (10 µl) and Ladder (3 µl, Page Ruler™ pre-stained protein ladder product# 26616) were loaded onto wells in 10% NuPAGE Bis-Tris gel and separated at 200 voltage for 35 minutes in NuPAGE MES SDS Running buffer (20x). Separated proteins were transferred to a PVDF membrane by stacking it on top of the gel in between filters and sponges soaked in a mixed transfer buffer (25 ml 20x NuPAGE Transfer buffer, 50 ml methanol and deionized

water up to 500 ml). The transfer was carried out in the mixed transfer buffer at 30 voltages for 90 minutes. The PVDF membrane was blocked with Odyssey blocking buffer (TBS) and 1xTBS (50/50) and incubated overnight with primary antibodies in a cool and dark place (Table 2). The day after the membrane was rinsed and incubated with secondary antibodies (anti-mouse and anti-rabbit from LI-COR) diluted 1:20.000 in 0.1% TBS/Tween and 0.02% SDS. The membrane was visualized with Odyssey from LI-COR Biotechnology (Cambridge, United Kingdom).

Table 2 Information on antibodies for Western blot

Antibody	Clone	Species	Isotype	Dilution	Company Order number
Actin	ACTNOS (C4)	Mouse	IgG1	1:2500	abcam ab3280
CD44	515	Mouse	IgG1	1:1000	BD Pharmingen 550988
CD324 (E-Cad)	36/E-Cadherin	Mouse	IgG2a, κ	1:1000	BD Bioscience 610182
CD325 (N-Cad)	32/N-Cadherin	Mouse	IgG1	1:1000	BD Bioscience 610921
CK5/6	D5/16B4	Mouse	IgG1	1:1000	Zymed 180267
CK14	Polyclonal	Rabbit	IgG	1:1000	abcam ab15461
CK18	DC10	Mouse	IgG1	1:1000	DAKO M7010
CK19	EP15801	Rabbit	IgG	1:1000	abcam ab52625
EpCAM	Polyclonal	Rabbit	IgG	1:2000	abcam ab71916
Thy-1	Polyclonal	Rabbit	IgG	1:500	Cell signaling CS#9798

3.9 3D culture

Three dimensional (3D) culture was carried out in reconstituted basement membrane (rBM), commercially known as Matrigel™ which is growth factor reduced (BD Biosciences). Matrigel was thawed and kept on ice in a liquefied state. Matrigel (300 µl) was added to 1000 D492 cells (sorted ALDH high, low, medium) and 80.000 BRENCs and mixed carefully to avoid bubbles forming. The Matrigel was divided to three wells (100 µl each) for each sorted cell type. The same amount of Matrigel was added to 20.000 of ALDH sorted D492 cells and divided to three wells. The procedure was done twice, ending up with 12 wells for each sorted cell type, six with D492 and BRENCs and six

with D492 without BRENCs. The wells were in a 96 well plate. The plate was stored in an incubator at 37°C for 30 minutes until the Matrigel solidified. Warm EGM5 medium (150 µl) was added to three wells of D492 and BRENCs and three without BRENCs for each sorted cell type. H14 medium was added to the other wells. The plate was kept in 37°C, 5% CO_2 and medium was changed every 3 days. Colony formation in the 3D culture was photographed after 8 days and before terminating the culture on day 16.

3.9.1 Isolation of colonies from 3D culture

Isolation of colonies from 3D cultures (Matrigel) was done as previously described by Bissell et al. (62). The entire process was carried out on ice. The medium was aspirated from the wells and washed with ice-cold PBS before adding 150 µl of PBS-EDTA (5 mM EDTA in PBS) to each well. By using a cut off pipette tip the gel was detached from the bottom of the well and the gel was transferred to a tube. The wells were rinsed with more PBS-EDTA to make sure that the entire content had been transferred. The tubes were filled with more PBS-EDTA (up to 1 ml) and gently shaken on ice for 30 minutes or until the gel had dissolved. The tubes were centrifuged in a cooled centrifuge (4°C) at 1000 RPM for 1-2 minutes. Majority of supernatant was aspirated and colonies re-suspended in the remaining supernatant. The colony mixture was placed in chambers on a chamber slide, 15 µl for each chamber and air-dried.

3.9.2 Fixation

Once dry, the colonies were fixed with 4% paraformaldehyde at RT for 10 minutes, 100 µL for each chamber. Fixation was stopped by aspirating the paraformaldehyde and adding 200 µl of PBS-glycine (100 mM glycine in PBS). The fixed colonies were eventually washed with PBS and stored in PBS at 4°C for 4 days.

3.9.3 Cytokeratin immunofluorescence (IF) staining of colonies

Before IF staining, colonies were washed three times with PBS-glycine for 10 minutes. The surface of the chambers was blocked with 100 µl of IF blocking solution (10% anti-goat serum, 1% goat $F(ab')_2$ anti-mouse immunoglobulin G in IF buffer (0.1% BSA, 0.2% Triton X-100, 0.05 Tween 20 in PBS)). Colonies were stained with primary antibodies for cytokeratin diluted in 100 µl of IF buffer overnight at 4°C (Table 3). After incubation the colonies were washed three times for 20 minutes with IF buffer and stained with secondary antibodies diluted 1:1000 in 100 µl of IF buffer for 45 minutes. Colonies were then after washed with IF buffer for 20 minutes and twice with PBS for 10 minutes. Counterstaining of nuclei was carried out with DAPI (1 µl/ml, diluted 1:5000 in IF buffer) for 15 minutes. Colonies were washed with PBS for 10 minutes and mounted with hard-set mounting medium.

3.10 Fixation and IF staining of monolayer cells

ALDH sorted cells were stained in a monolayer. The cells were originally seeded on an untreated polystyrene 12 well plate (Falcon), 1000 cells per well and cultured for 10 days on EGM5. Cells were

trypsinized and transferred to a chamber slide. Cells were fixed with ice cold methanol for 10 minutes and blocked with 10% FBS. Cells were incubated overnight at 4°C with primary antibodies (Table 3) diluted in IF buffer. The day after, the cells were rinsed with PBS and incubated for 30 minutes at RT with secondary antibodies diluted 1:1000 in IF buffer. Following incubation cells were rinsed with PBS and incubated for 15 minutes with DAPI for counterstaining of nuclei. Cells were again rinsed with PBS and finally deionized water. After the slide had dried it was mounted with hard-set mounting medium.

Table 3 Information on antibodies for IF staining of monolayer cells and 3D culture

Antibody	Clone	Species	Isotype	Dilution	Company Order number	Secondary antibody
CK5/6	D5/16B4	Mouse	IgG1	1:50	Zymed 180267	Alexafluor 647 (blue) goat anti-mouse IgG1
CK8	35bH11	Mouse	IgG1	1:100	abcam ab9023	Alexafluor 647 (blue) goat anti-mouse IgG1
CK14	LL002	Mouse	IgG3	1:25	Novocastra NCL-L-LL002	Alexafluor 488 (green) goat anti-mouse IgG3
CK17	E3	Mouse	IgG2b	1:50	DAKO M7046	Alexafluor 546 (red) goat anti mouse IgG2b
CK18	DC10	Mouse	IgG1	1:50	DAKO M7010	Alexafluor 647 (blue) goat anti-mouse IgG1
CK19	A53-B/A2	Mouse	IgG2a	1:100	abcam ab7754	Alexafluor 488 (green) goat anti-mouse IgG2a

4 Results

4.1 Comparison of D492 and D492M in relations to ALDH activity and cell surface marker expression

D492 and D492M were stained for ALDH activity and surface markers CD24, CD29, CD44, CD90, CD117, CD133 and CD324. Data was analyzed using consistent gating definitions for both cell lines.

4.1.1 ALDH activity in D492 and D492M

ALDH activity was compared between D492 and D492M. The high ALDH activity population was defined according to a baseline observed in the negative control. For consistency, the same gate was used for every negative control and sample of D492 and D492M.

Comparison of ALDH activity in D492 and D492M shows a higher activity in D492 (Figure 5). The highest value for D492 is 2.12% vs. 0.52% in D492M. The average ALDH activity for both cell lines is shown in a histogram in Figure 6. According to Student´s T-test there is significant difference between ALDH activity in D492 and D492M with P value <0.05.

Combined histograms for ALDH stained populations of negative controls and samples display the difference between D492 and D492M (Figure 7A and B) and difference between controls and samples for each triplicate of D492 and D492M is shown in Figure 7C-F. The combined histograms illustrate marked instability in the negative control for D492. The negative controls for D492M appear less variable.

According to images of the cultures taken before execution of the Aldefluor assay, there seems not to be variability in the appearance of D492 and D492M from well to well (Supplementary figure 1).

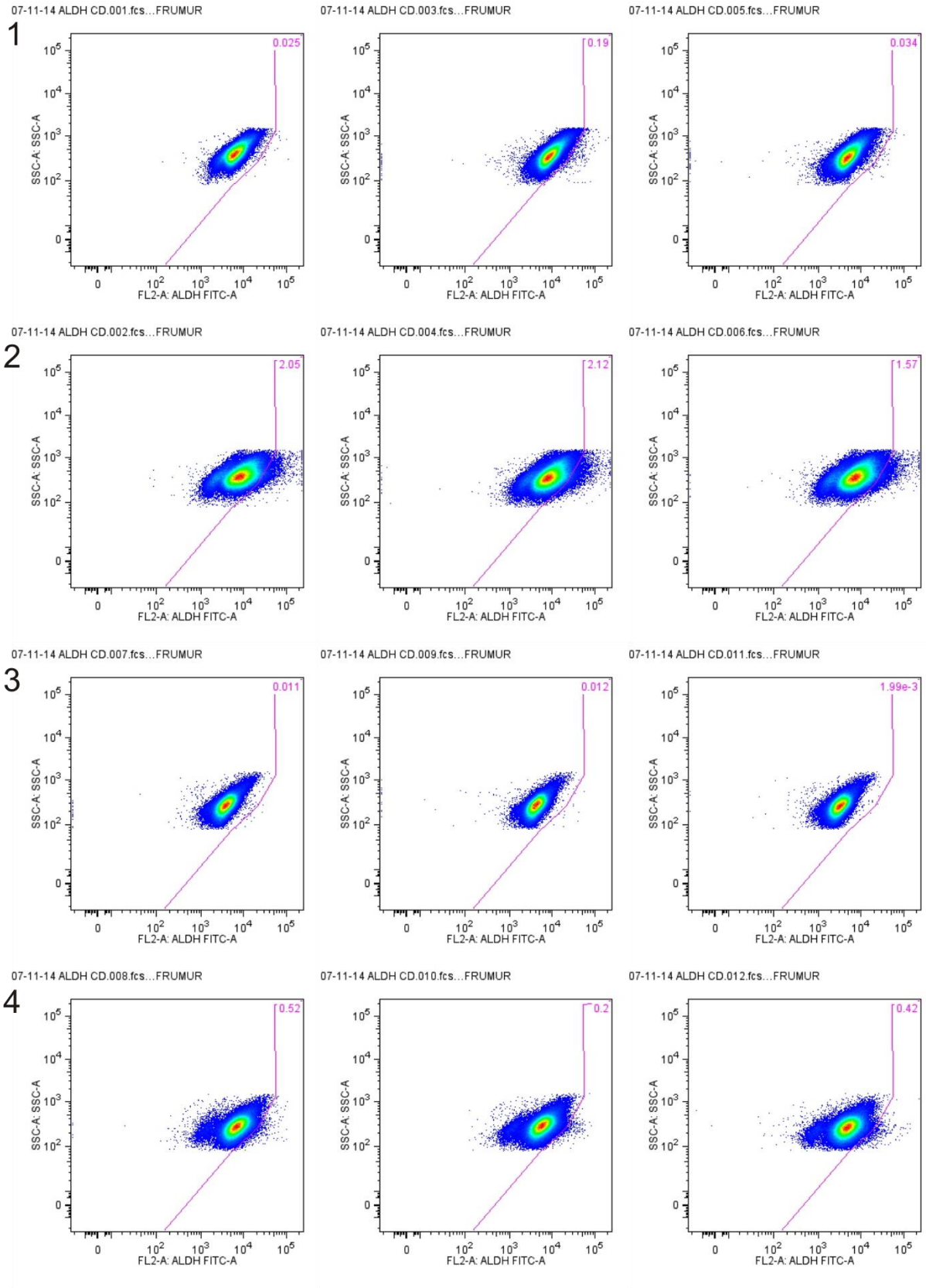


Figure 5 ALDH activity in D492 and D492M

Row 1: Negative controls of D492. **Row 2:** Samples of D492 **Row 3:** Negative controls of 492M. **Row 4:** Samples of D492M.

Each negative control and sample cultured in triplicate. The controls and samples are consecutively aligned and each sample corresponds to the control above it.

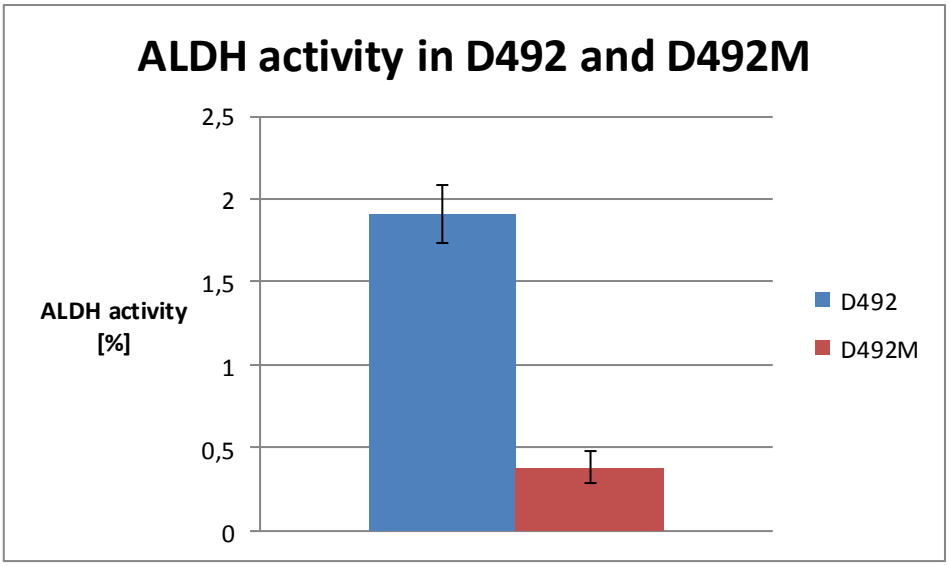


Figure 6 ALDH activity compared between D492 and D492M – histogram

The histogram represents average ALDH high activity in D492 and D492M. Significant difference is between D492 and D492M with P value < 0.05. Error bars are set according to the standard error of mean (SEM).

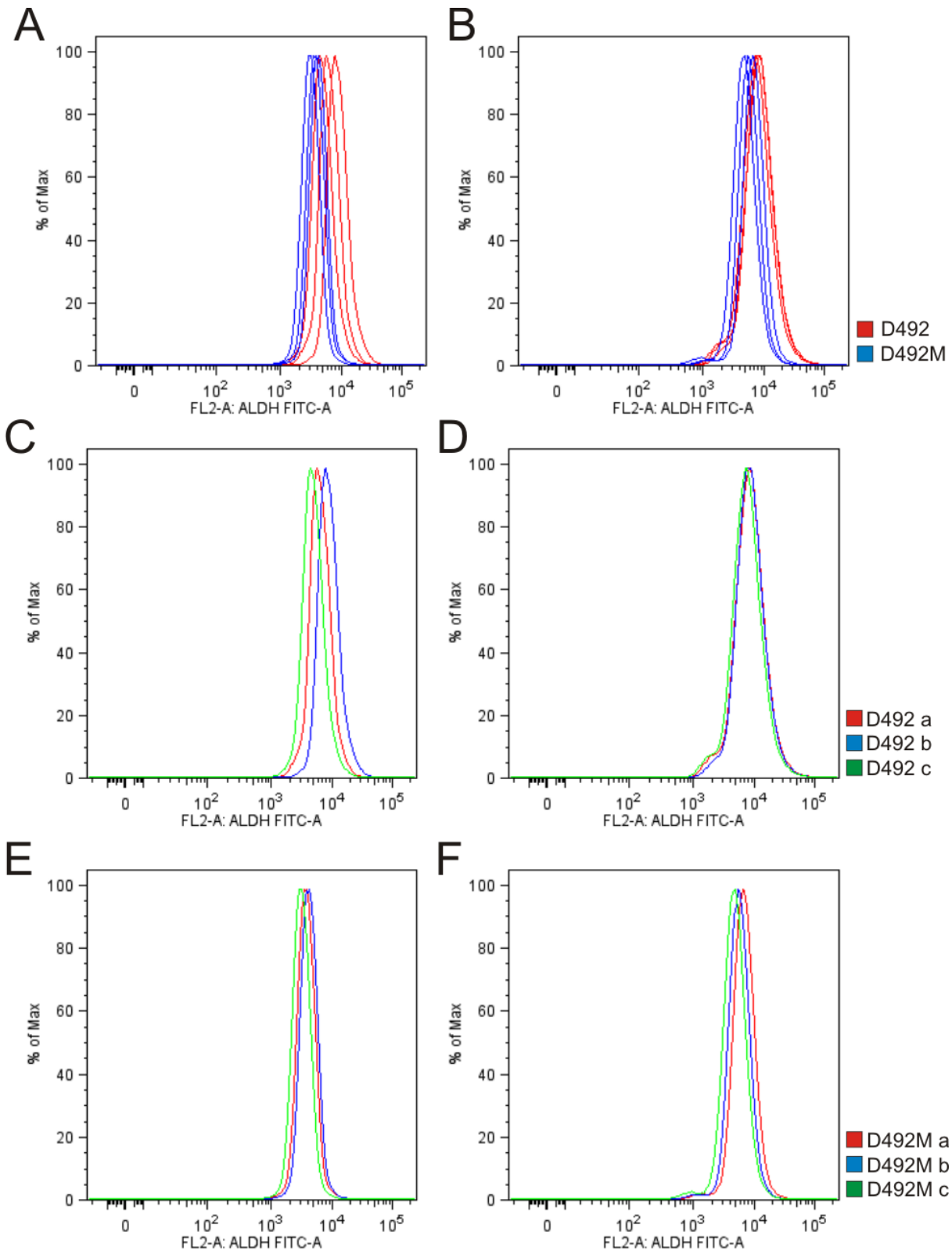


Figure 7 Comparison of ALDH activity in D492 and D492M – histogram curves combined

A: Negative controls for D492 and D492M combined. **B:** Samples of D492 and D492M combined. **C:** Negative controls of D492 combined. **D:** Samples of D492 combined. **E:** Negative controls of D492M combined. **F:** Samples of D492M combined.

Fluorescence ALDH measurement of both negative controls and samples (triplicate for D492 and D492M) displayed in a histogram curve and grouped for comparison of ALDH activity between D492 and D492M.

4.1.2 Expression of surface markers on cells with ALDH high and low activity

Selection of ALDH activity levels for definition of ALDH high and low population were set at the 5% (from 4.87 to 5.49%) extreme in D492 and D492M. Expression of the potential stem cell markers was evaluated in these high and low 5%, always including CD44 for a consistent comparison.

For markers CD24, CD90 (Thy-1), CD117 and CD29 against CD44, a similar pattern was observed in both cell lines (Figure 8 and 9). Generally ALDH high cells show a more intense marker staining than ALDH low cells.

CD133 displays two separate populations with respect to marker intensity in ALDH low cells for D492; a high and low populations. The low population has the same appearance as observed in ALDH high cells (Figure 8C and F). It seems that the low population shows a higher CD133 and CD44 expression in ALDH high cells while the high population disappears.

In D492M (Figure 8I and L) the same occurs for CD133; two populations are formed in ALDH low cells. The separation between populations is however not as clear as in D492 as the higher population in D492M contains fewer cells. A similar shift of the lower population is observed, with higher expression of markers in ALDH high compared to ALDH low.

Increased expression of ALDH high cells is also shown for markers CD324 (E-Cad) and CD44 in D492 (Figure 9C and F) and D492M (Figure 9I and L). However the appearance of the population is not exactly the same between ALDH high and low as seen for other markers.

The experiment was done in triplicate but only one selected analysis is displayed in the results, data for marker expression in all three measurements can be viewed in supplementary figures 2 -7.

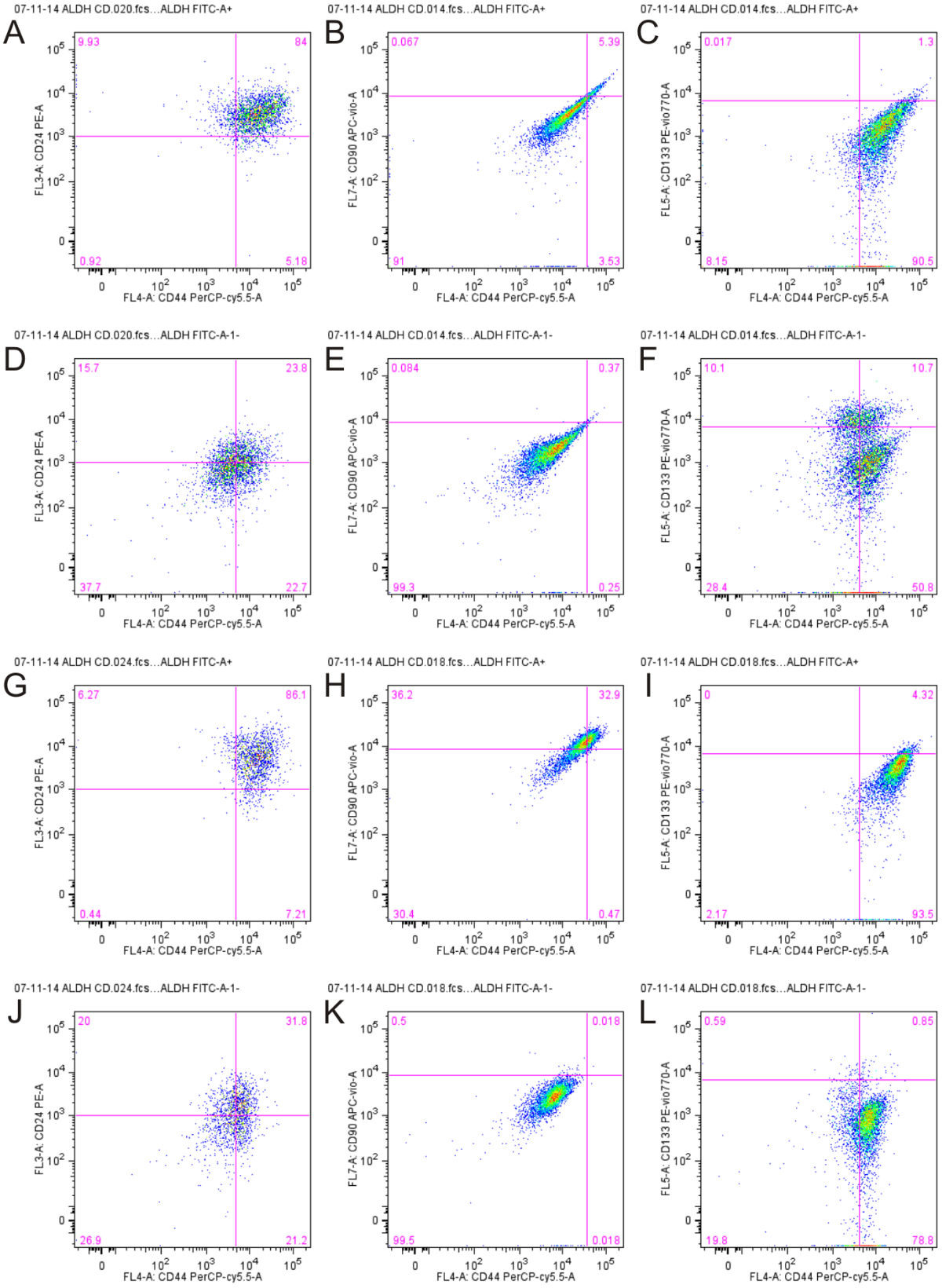


Figure 8 CD24, CD90 and CD133 expression in 5% ALDH high and low D492 and D492M cells

A: CD24 and CD44 in ALDH high D492. **B:** CD90 and CD44 in ALDH high D492. **C:** CD133 and CD44 in ALDH high D492.

D: CD24 and CD44 in ALDH low D492. **E:** CD90 and CD44 in ALDH low D492. **F:** CD133 and CD44 in ALDH low D492.

G: CD24 and CD44 in ALDH high D492M. **H:** CD90 and CD44 in ALDH high D492M. **I:** CD133 and CD44 in ALDH high D492M. **J:** CD24 and CD44 in ALDH low D492M. **K:** CD90 and CD44 in ALDH low D492M. **L:** CD133 and CD44 in ALDH low D492M.

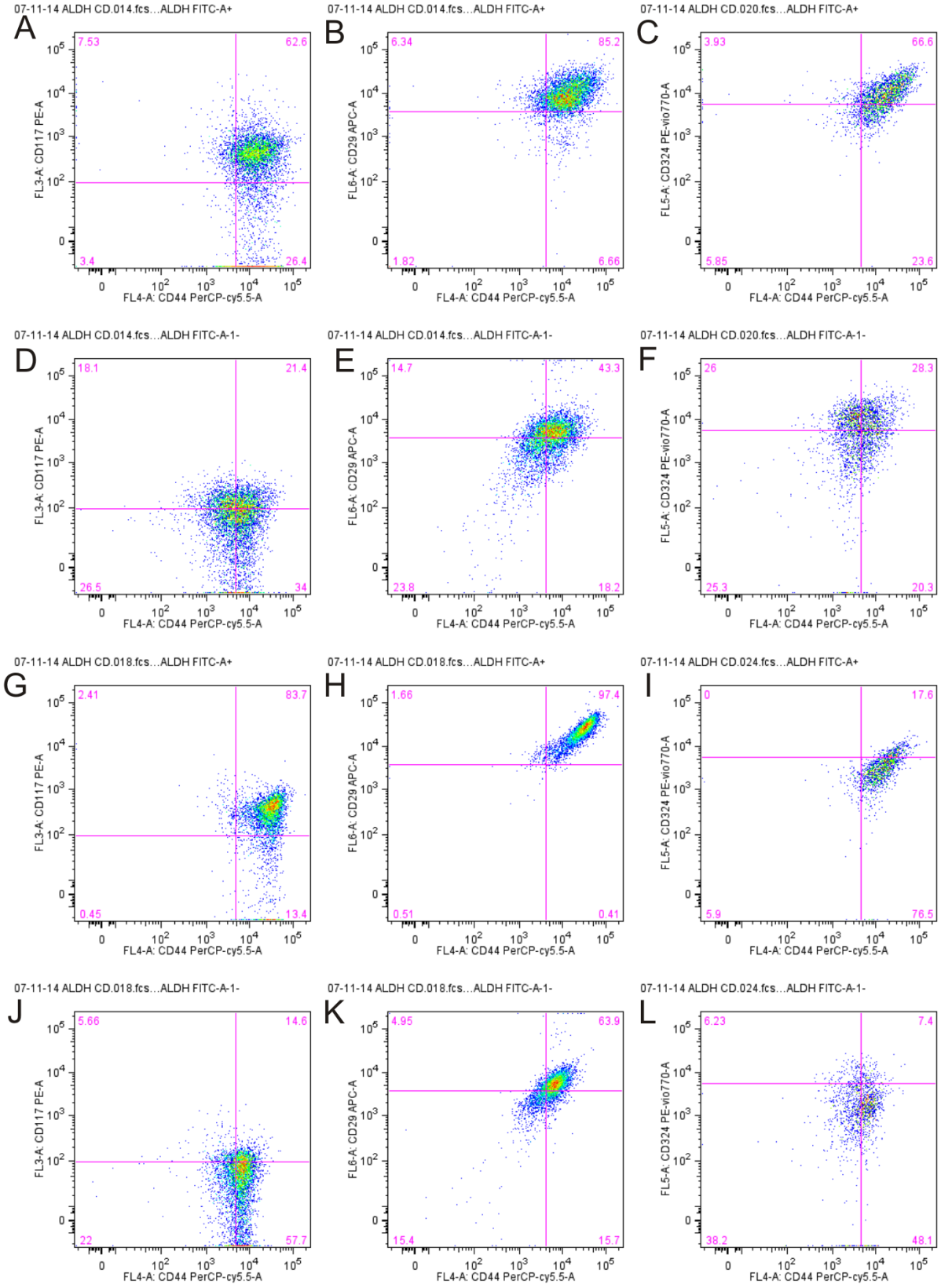


Figure 9 CD117, CD29 and CD324 expression in 5% ALDH high and low D492 and D492M cells

A: CD117 and CD44 in ALDH high D492. **B:** CD29 and CD44 in ALDH high D492. **C:** CD324 and CD44 in ALDH high D492. **D:** CD117 and CD44 in ALDH low D492. **E:** CD29 and CD44 in ALDH low D492. **F:** CD324 and CD44 in ALDH low D492. **G:** CD117 and CD44 in ALDH high D492M. **H:** CD29 and CD44 in ALDH high D492M. **I:** CD324 and CD44 in ALDH high D492M. **J:** CD117 and CD44 in ALDH low D492M. **K:** CD29 and CD44 in ALDH low D492M. **L:** CD324 and CD44 in ALDH low D492M.

4.1.3 Comparison of marker expression in D492 and D492M

Comparison of marker expression in D492 and D492M is shown in Figure 10 with combined histogram curves for every sample. Expression of CD24 is increased in D492M compared to D492; the expression however varies between samples (Figure 10A). CD90 (Thy-1) is increased in D492M compared to D492 (Figure 10C). CD324 (E-Cad) expression is clearly increased in D492 compared to D492M (Figure 10G). The expression of other markers is similar between D492 and D492M (Figure 10B, D, E and F).

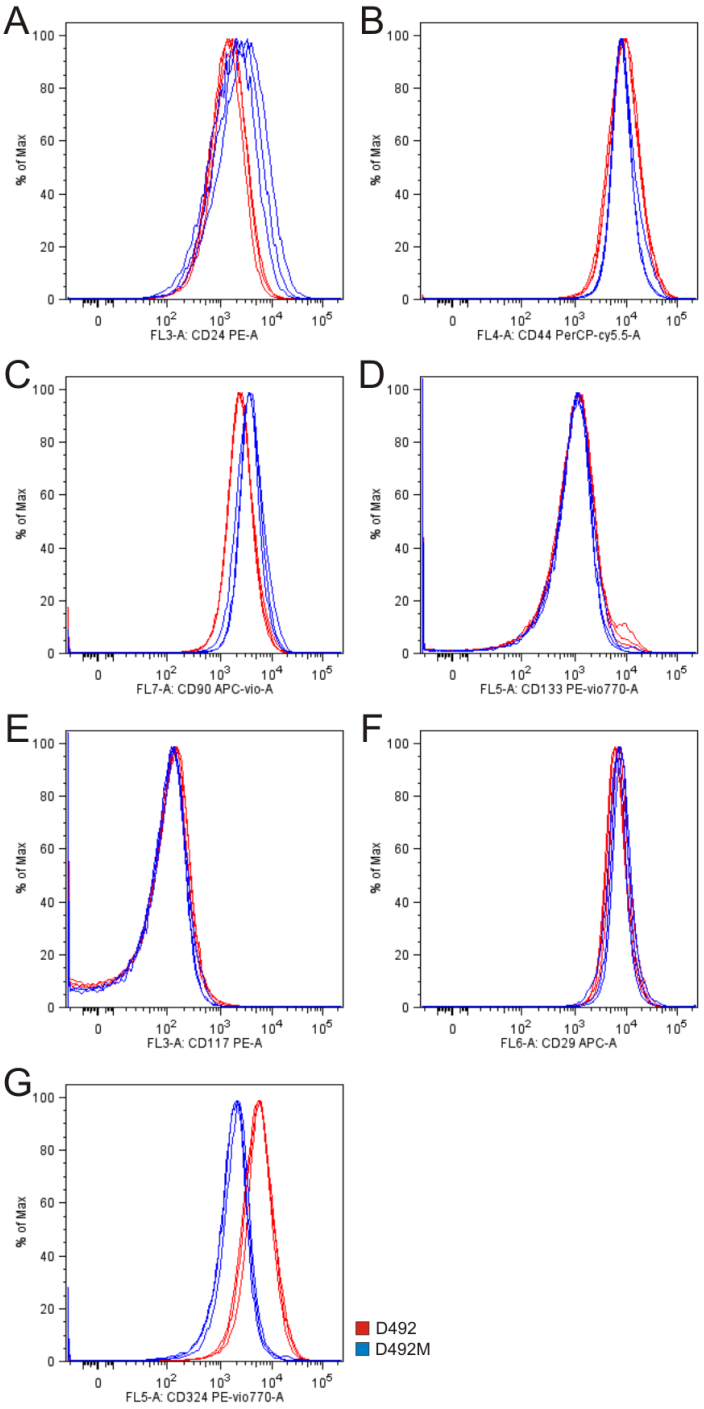


Figure 10 Comparison of marker expression in D492 and D492M – Histogram curves combined

A: CD24. **B:** CD44. **C:** CD90 (Thy-1). **D:** CD133. **E:** CD117. **F:** CD29. **G:** CD324 (E-Cad).

4.1.4 CD24 and CD44 expression in D492 and D492M

CD24 expression increases in D492M as seen in the higher gate of the dot plot in Figure 11A and B. CD24 low expression is consistent between cell lines, seen in the lower gate. The CD44 expressing population is wider in shape in the dot plot for D492. A larger portion of the D492 population falls in the right quadrats compared to D492M.

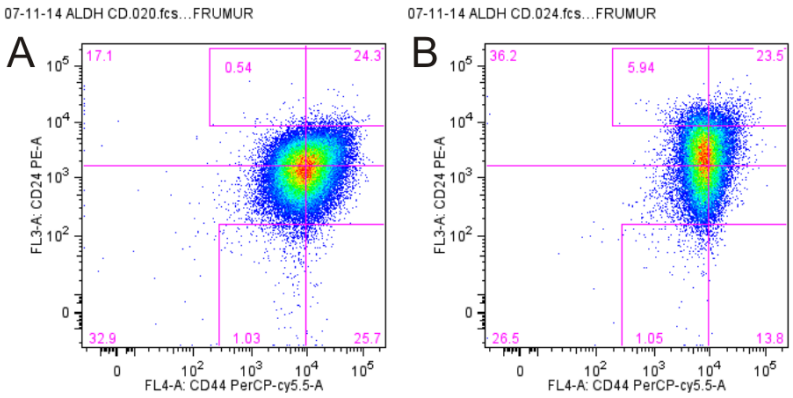


Figure 11 Expression of CD24 and CD44 in D492 and D492M

A: D492 **B:** D492M.

Expression of CD24 increases in D492M while CD44 expression decreases slightly compared to D492.

4.1.5 Different SSC properties of D492 and D492M

D492 and D492M have obvious different SSC properties as seen in Figure 12. The D492M population has a lower SSC than D492.

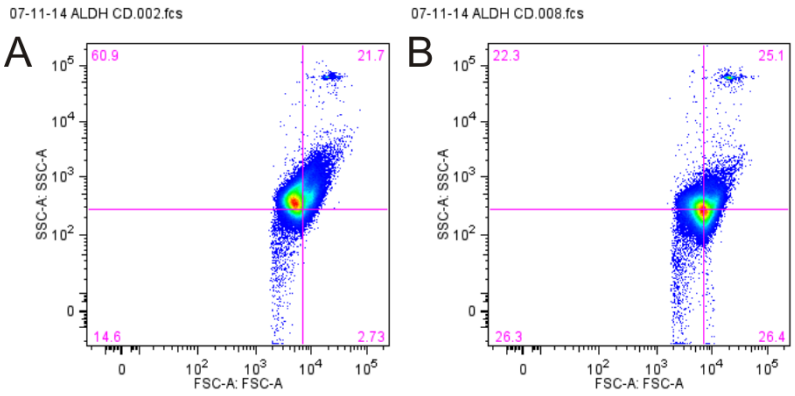


Figure 12 Differences in SSC of D492 and D492M

A: D492. **B:** D492M

4.2 ALDH activity in D492 cells with different passage number

ALDH activity was measured in four different cell passages (30, 41, 50 and 60) of D492, in triplicate for every passage. ALDH high activity ranged from 1.26% to 3.58% according to a gating strategy using negative controls to define highly expressing cells (Figure 13).

A histogram of the average ALDH activity for each passage is shown in Figure 14. Passage 60 has the lowest ALDH activity and passage 41 the highest. However according to one-way non-parametric ANOVA test there is no significant difference between the ALDH activities in different cell passages (P>0.05).

Histograms for ALDH of the different passages of D492 reflect only minor differences (Figure 15B). However, there is less consistency in the negative controls since the activity in cells of passage 50 is clearly decreased (Figure 15A).

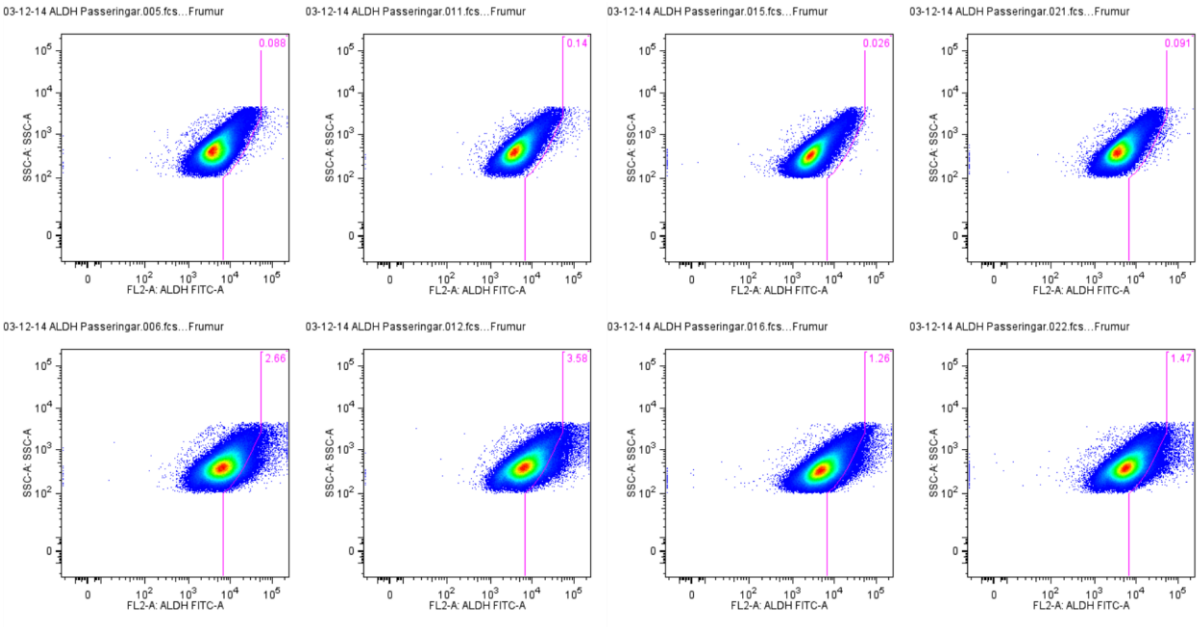


Figure 13 ALDH activity in different cell passages of D492

Top row: Negative control (DEAB) for passages 30, 41, 50 and 60. **Bottom row:** Samples of passages 30, 41, 50 and 60, corresponding to negative controls in the top row. Only one of three controls and samples are shown for each passage with the highest and lowest activity included. Analysis of other controls and samples are shown in supplementary figure 11 and 12 in appendix.

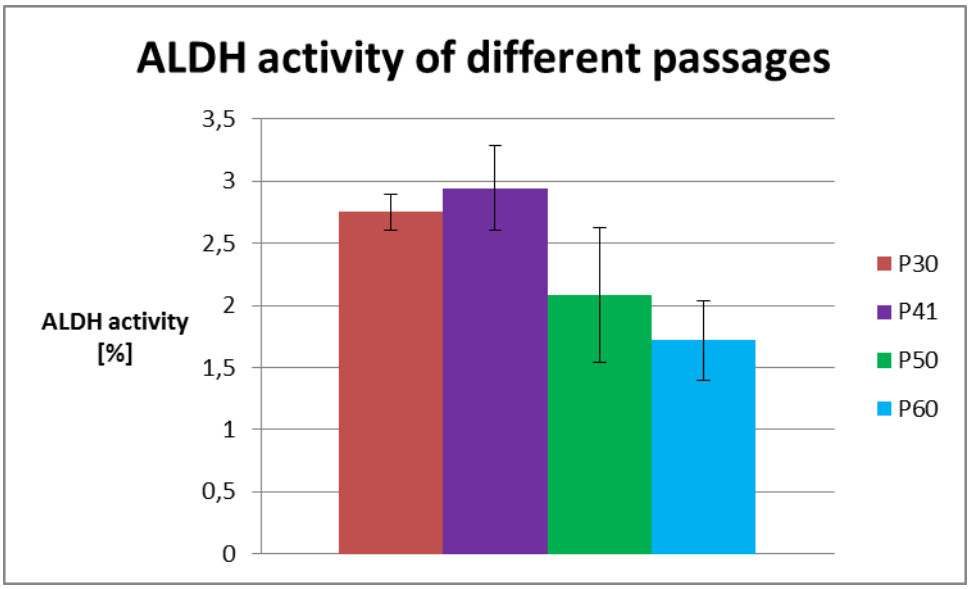


Figure 14 Histogram of ALDH activity in different passages

Histogram displaying the average ALDH activity for each passage. Statistical analysis reveals no difference between passages as P value is greater than 0.05. Error bars are set according to SEM

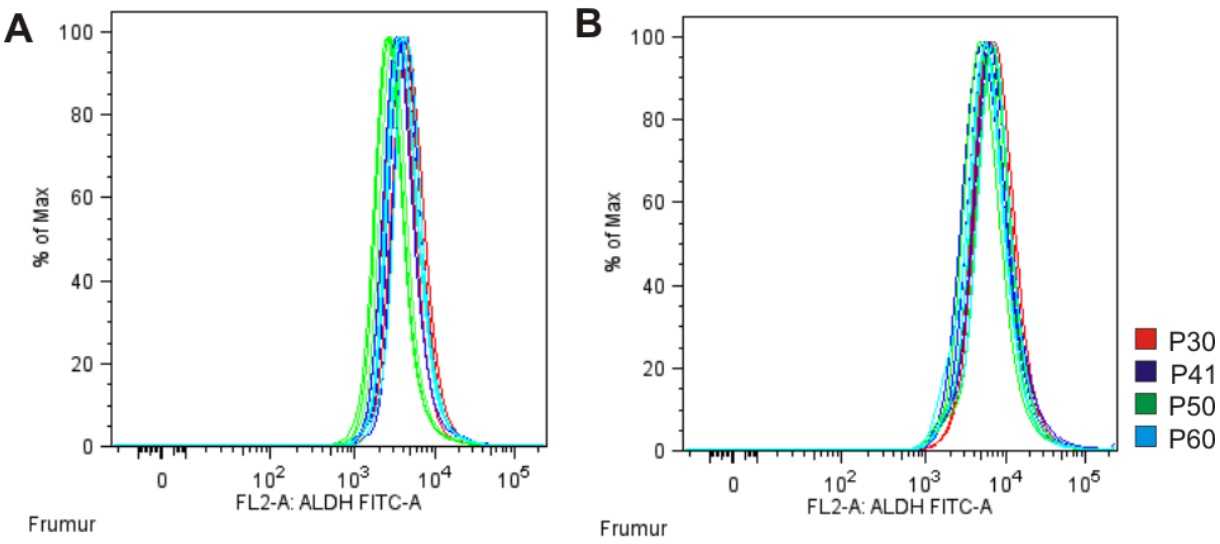


Figure 15 ALDH activity in different cell passages of D492 – histogram curves combined

A: Negative controls. **B:** Samples.

Histogram for all ALDH fluorescence measurements of negative controls and the samples (three for every passage) combined for a visual comparison.

In monolayer cell cultures from different passages, appear morphologically similar as seen in supplementary figure 10.

4.3 Sorting of ALDH high, medium and low cells in FACSAria™ II

D492 stained for ALDH activity was sorted in FACSAria into three distinct populations; ALDH high, medium and low. The ALDH high fraction was defined as 10% of cells with highest activity and low was correspondingly defined as the 10% with lowest activity. The medium fractions contained 48% of the cells distributed around the mean.

The number of cells sorted for each type and number of viable cells after is shown in Table 4. The original number of cells brought to sorting was 20 million, however not all subjected to sorting due to a time factor.

During sorting, the cells seemed to lose the fluorescence signal progressively probably due to elevated temperature. For instance, the fraction of cells in the original ALDH high gate (10%) dropped down to 3% at the end of the sorting procedure if no actions were taken. In order to account for this, the gates were moved at regular time intervals to maintain the selection at 10%.

Table 4 Number of cell sorted in FACSAria™ II and viable cells after sorting

	ALDH high cells	ALDH medium cells	ALDH low cells
Sorted in FACSAria™ II	787.665	956.385	1.113.523
Viable after sorting	244.000	220.000	272.000

4.4 Post-sorting ALDH levels in D492 compared to unsorted cells

ALDH high, medium and low cells sorted in FACSAria were seeded at different density (1000, 2000 and 5000 cells in a 12 well plate) and cultured for 10 days on EGM5 medium. ALDH activity was measured in these cell types and compared using gates adjusted by using the negative control. The most variance was seen when 5000 cells were seeded. The results for that density are mostly informative and relevant. Results for 1000 and 2000 cells are shown as supplementary data in appendix (pages 90-92).

4.4.1 ALDH activity – 5000 cells seeded

A striking variance between ALDH sorted populations was seen when 5000 cells were seeded. The highest activity is maintained in ALDH high cells and the lowest in the sorted ALDH low cells (Figure 16).

However differences are seen in the negative controls when ALDH histograms are combined. Controls for ALDH medium and low cells are in agreement while controls for ALDH high fraction displays higher ALDH fluorescence (Figure 17A). ALDH fluorescence for the samples shows a corresponding difference, with ALDH high fraction displaying the brightest signal (Figure 17B).

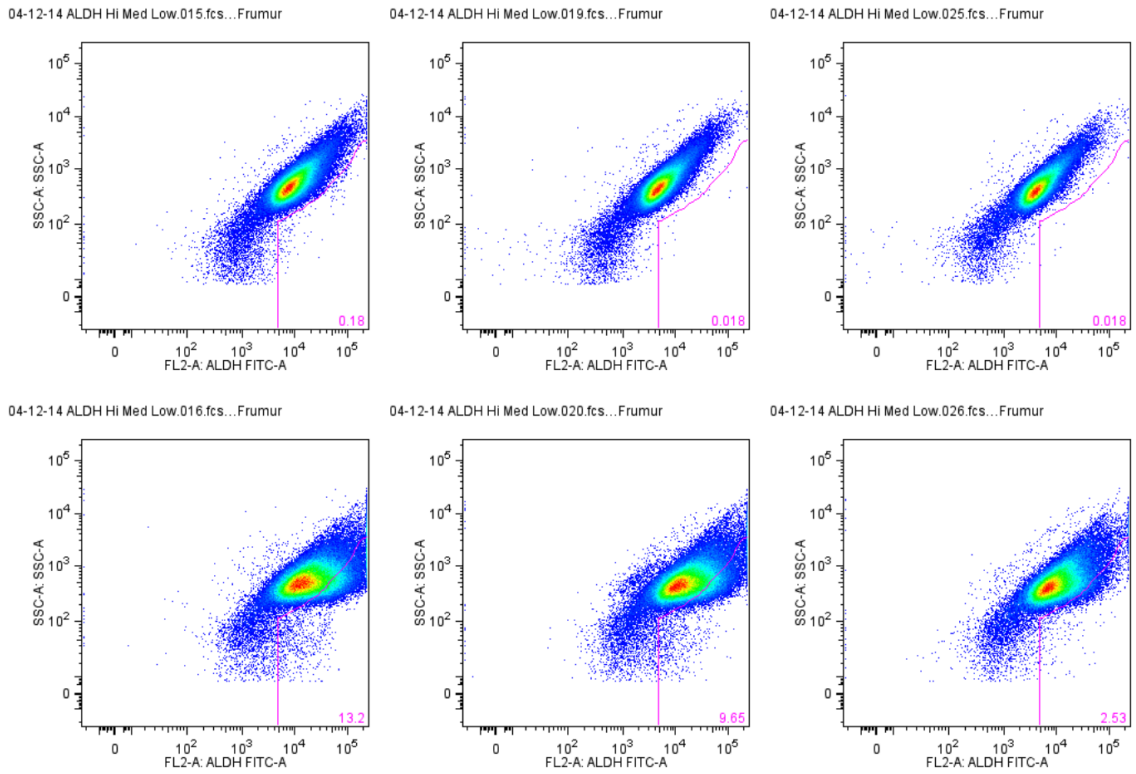


Figure 16 FACS analysis of ALDH activity in ALDH sorted high, medium and low cells – 5000

ALDH sorted populations, from cultures of 5000 cells seeded. **Top row:** Negative control for ALDH high, medium and low cells. **Bottom row:** Samples for ALDH high, medium and low cells matched to the negative controls above.

Analysis was performed in triplicate; other results are shown in supplementary figures 13-15.

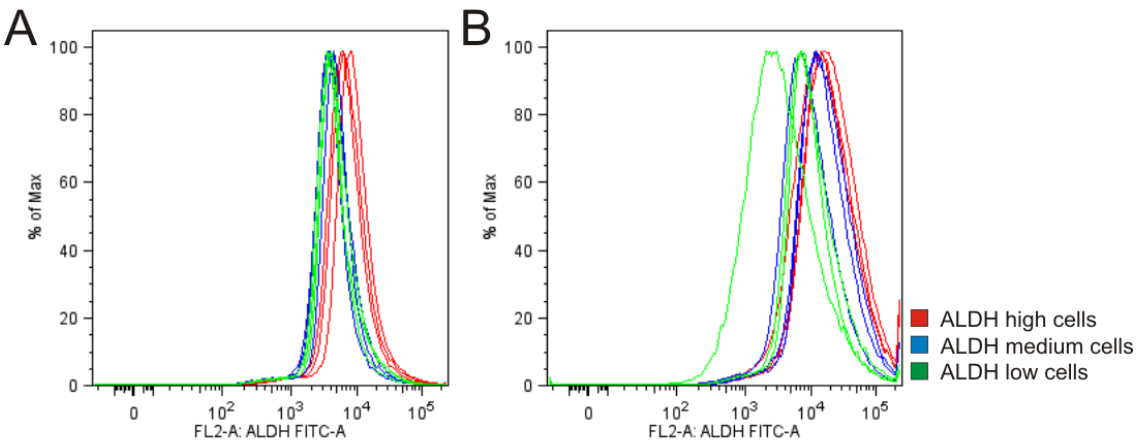


Figure 17 ALDH activity in 5000 seeded cells – histogram curves combined

A: ALDH Histograms for negative controls. **B:** ALDH histograms for samples.

4.4.2 Comparison of 1000, 2000 and 5000 cells seeded.

Comparison of 1000, 2000 and 5000 cells seeded show that the ALDH activity is higher with 5000 cells seeded for both ALDH high and medium fractions (Figure 18B and D). For ALDH low cells the distinction between number of cells seeded is less (Figure 18F), with the exception of one aberrant sample where 5000 seeded cells have a lower fluorescence signal. Histogram for controls is similar but do not line perfectly together (Figure 18A, C and E).

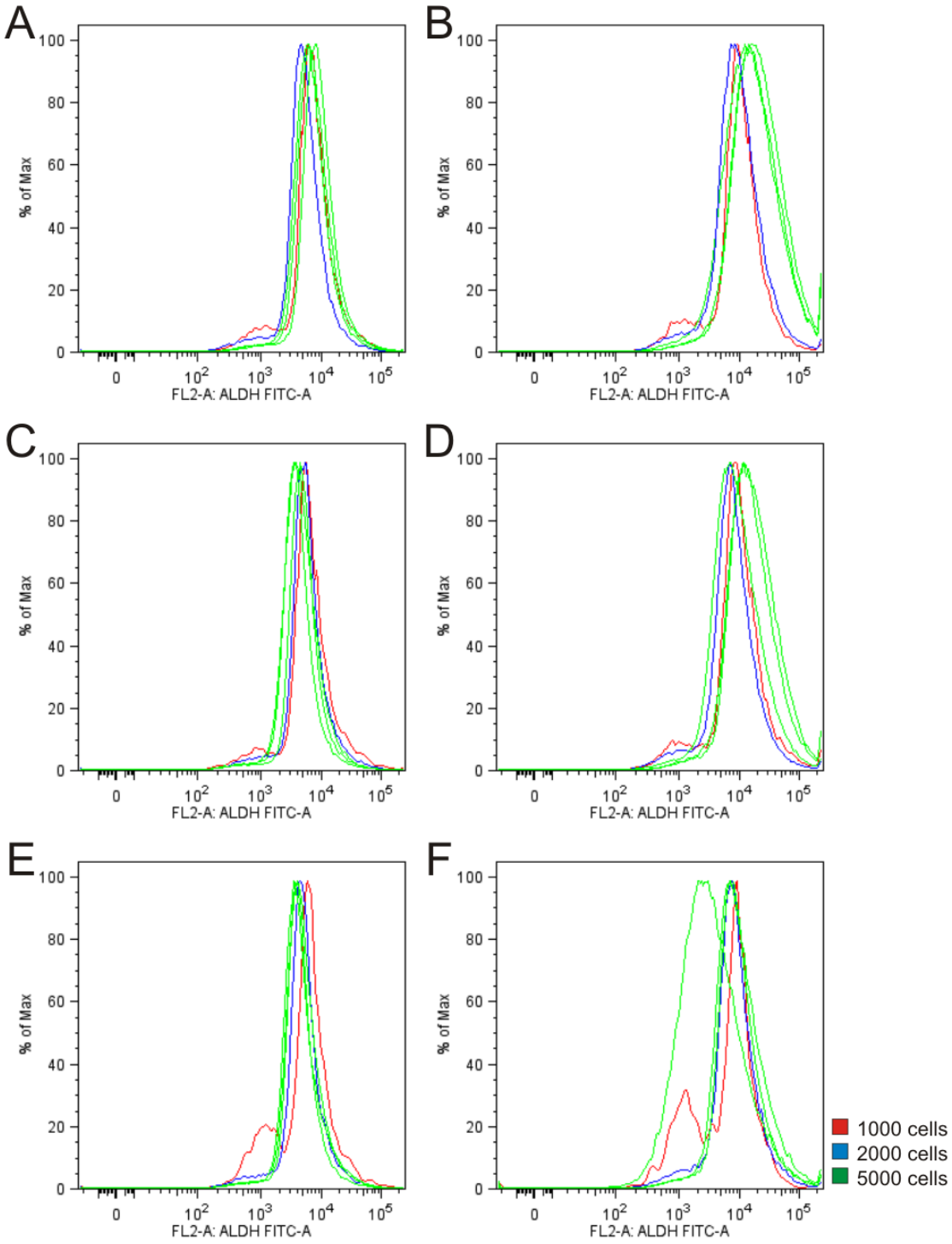


Figure 18 ALDH activity in cultures from ALDH sorted fractions – histogram curves combined

A: Negative controls for ALDH high cells. **B:** Samples for ALDH high cells. **C:** Negative controls for ALDH medium cells. **D:** Samples for ALDH medium cells. **E:** Negative controls for ALDH low cells. **F:** Samples for ALDH low cells.

4.4.3 ALDH activity in cultures from ALDH sorted high, medium and low cell fractions compared to ALDH activity prior to sorting

Comparison of ALDH activity in sorted cells and unsorted is shown in Figure 19. Sorted cells were grown in monolayer culture for 10 days before execution of Aldefluor assay. Unsorted cells did not go through FACSAria nor were they subjected to further culture, the Aldefluor assay was performed on the day of sorting.

Unsorted cells show lower levels of ALDH fluorescence in both negative control and sample (Figure 19A and B). Figure 19C and D show that the histogram for unsorted cells falls between ALDH sorted high and low cells. The unsorted cell histogram is more or less in the same range as ALDH low cells.

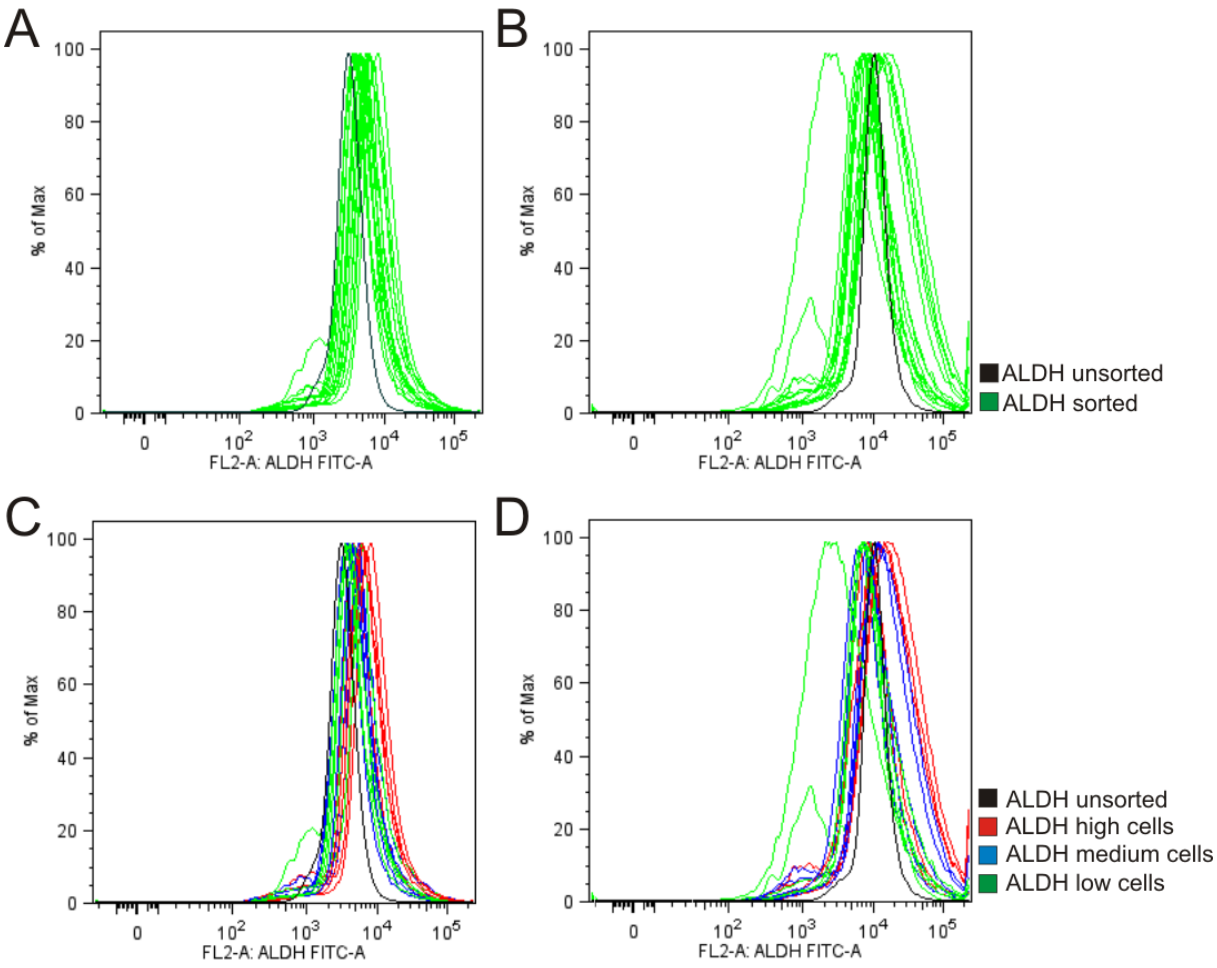


Figure 19 ALDH sorted high, medium and low cells compared to ALDH unsorted cells

A: Negative controls for all ALDH sorted and unsorted cells. **B:** Samples for all ALDH sorted and unsorted cells. **C:** Negative controls for ALDH sorted high, medium and low cells, histograms combined and compared to ALDH unsorted. **D:** Samples of ALDH sorted high, medium and low cells, histograms combined and compared to ALDH unsorted.

4.4.4 Indicator of proliferation and altered morphology in ALDH sorted cells

In order to evaluate cell proliferation in ALDH fractions, the cells were counted manually with hemocytometer prior to the Aldefluor assay and during FACS. The results are listed in Table 5.

Table 5 Counting of ALDH high, medium and low cells

Total number of cells counted both manually and during FACS analysis.

	ALDH high cells		ALDH medium cells		ALDH low cells	
Cells seeded	Manual	FACS	Manual	FACS	Manual	FACS
1000	22.400	20.221	18.000	23.950	10.000	5.377
2000	80.000	79.823	76.000	76.343	56.000	57.099
5000 a	172.000	230.795	210.000	232.873	204.000	171.721
5000 b	186.000	229.438	176.000	234.395	216.000	188.471
5000 c	278.000	284.400	232.000	276.124	172.000	191.565

Phase contrast light microscopy images taken of ALDH high, medium and low cells show the morphology of the cells and the approximate proliferation in 10 days of culture (Figure 20).

Images taken at a greater magnification (100x) show abnormalities including large and multinucleated cells in all sorted cell fractions when 1000 cells were seeded (Figure 21). Similar cells are also seen at other seeding densities (Figure 20). The morphology appears altered when compared to cells that have not undergone sorting (Supplementary figures 1 and 10).

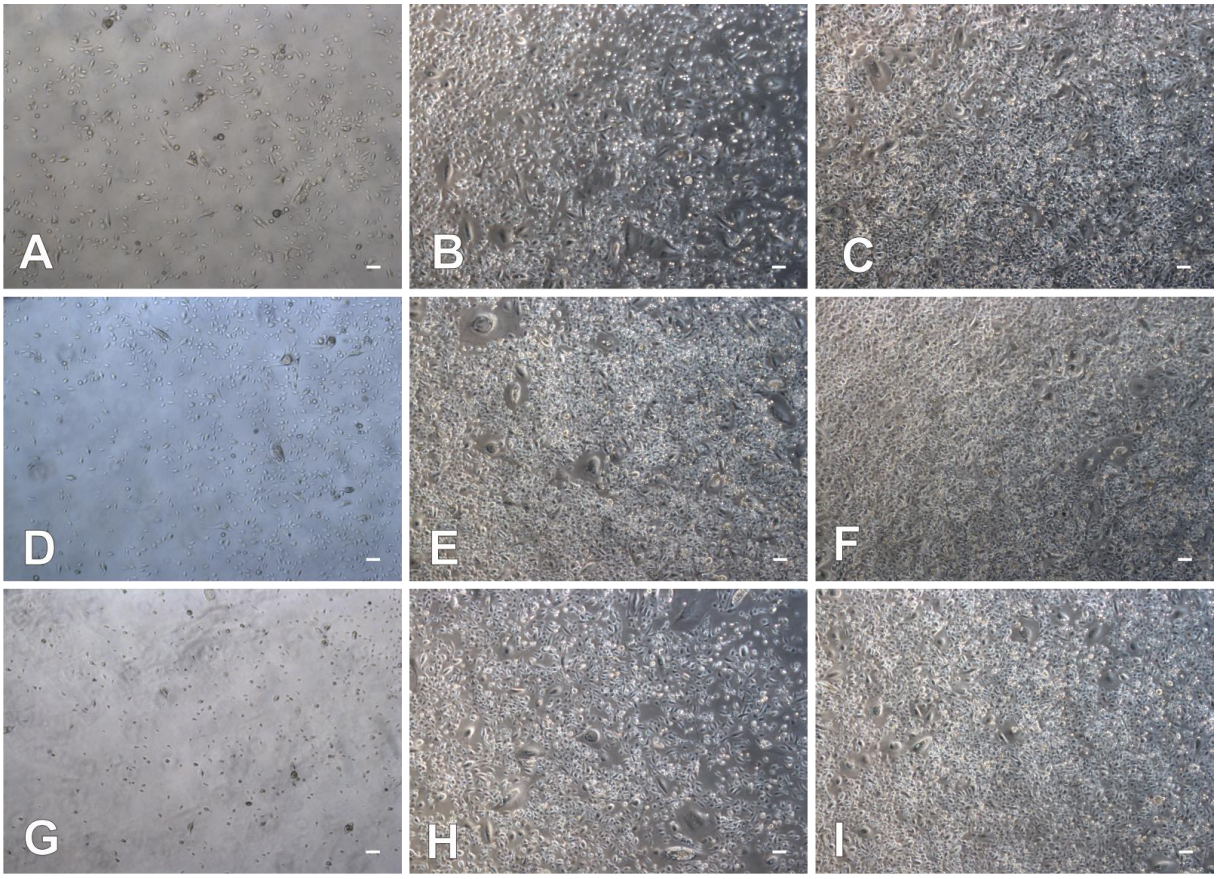


Figure 20 ALDH high, medium and low cells after 10 days in monolayer culture

A-C: ALDH high cells, 1000 (A), 2000 (B), 5000 (C) cells seeded. **D-F**: ALDH medium cells, 1000 (D), 2000 (E), 5000 (F) cells seeded. **G-I**: ALDH low cells, 1000 (G), 2000 (H), 5000 (I) cells seeded. Phase contrast light microscopy, original magnification 50x, bar – 100 µm.

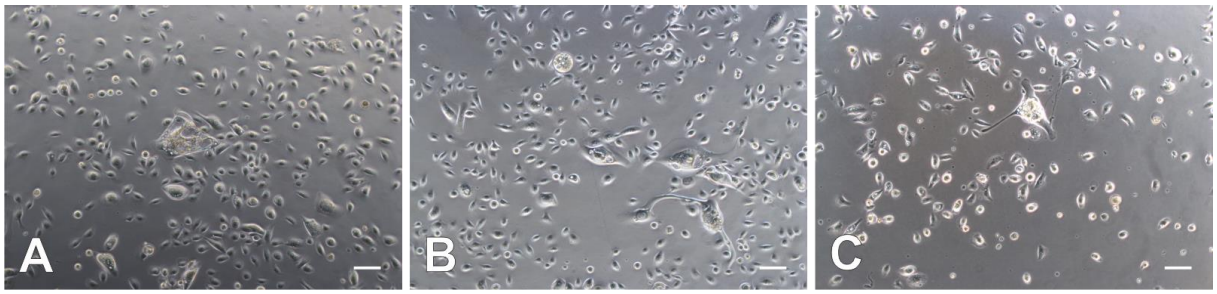


Figure 21 ALDH sorted high, medium, low cells in monolayer culture; 1000 cells seeded

A: ALDH high cells. **B:** ALDH medium cells. **C:** ALDH low cells. Large and multinucleated cells can be an indicator of stress in culture. Phase contrast light microscopy, original magnification 100x, bar – 100 µm.

4.5 Western blot for ALDH sorted fractions

Western blot was performed on the isolated proteins from ALDH high, medium and low cells sorted in FACSAria. Expression of proteins; Actin, E-Cadherin (E-Cad), EpCAM, Thy-1, CK14, CK5/6 and CK19 was observed. No results were obtained for CK18, CD44 and N-Cadherin. Original western blot data is shown in supplementary figures 20 and 21.

4.5.1 Actin, E-Cadherin and EpCAM

Western blot of actin, E-Cad and EpCAM was performed in duplicate for proteins from ALDH high (H1/H2), medium (M1/M2) and low (L1/L2) cells. Bands formed for each protein from all cell types are seen in Figure 22. Actin, which was equally expressed in every sorted type, was used to normalize the levels of E-Cad and EpCAM when quantified. Quantification data (Figure 23) shows equal expression of E-Cad in ALDH high and medium cells but lower in ALDH low cells. EpCAM is similar between ALDH high and low but highest for ALDH medium cells. Western blot for this protein expression was only done once and in duplicate only and therefore statistical analysis cannot be performed.

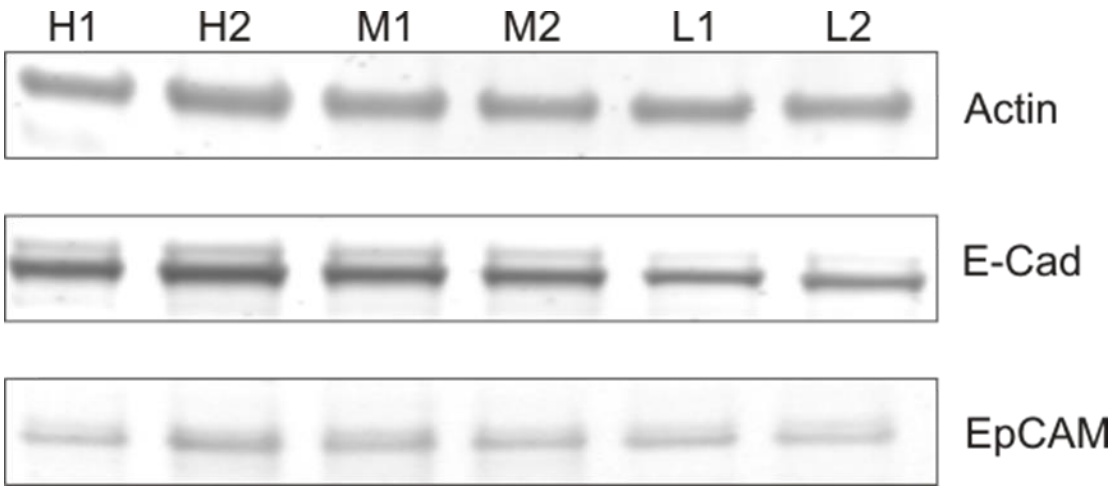


Figure 22 Western blot of Actin, E-Cadherin and EpCAM.

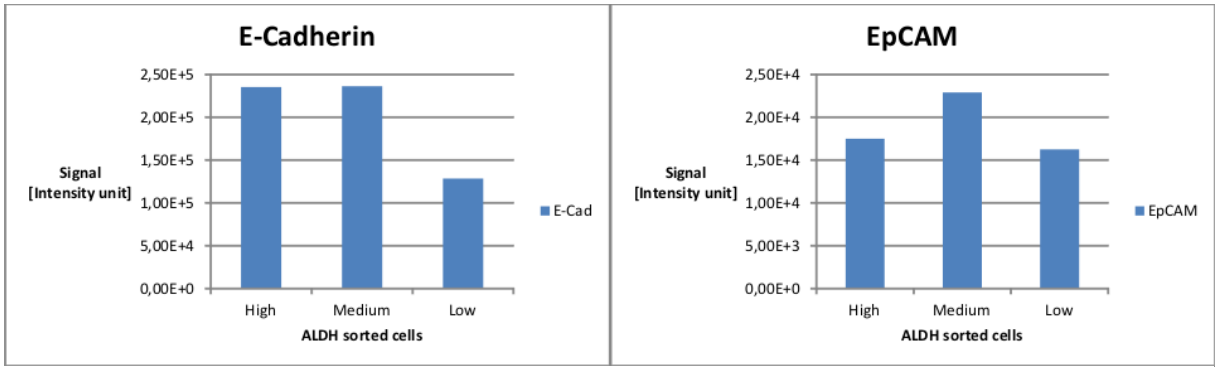


Figure 23 Western blot quantification data for expression of E-Cadherin and EpCAM

Histogram represents average protein expression for each sorted cell type (duplicate).

4.5.2 Actin, Thy-1, CK14, CK5/6 and CK19

Western blot of actin, Thy-1, CK14, CK5/6 and CK19 was performed using the same protein isolations as before from cell cultures of sorted cell fractions. As expression levels were similar between 1 and 2 only high (H1), medium (M1) and low (L1) proteins were used. The western blots are shown in Figure 24. Only minor differences in expression levels were seen in ALDH sorted fractions for CK14, 5/6 and 19 (Figure 25). The most robust difference between ALDH high, medium and low cells was seen for Thy-1 expression. Thy-1 expression was similar for ALDH high and medium but non-detectable in ALDH low cells. However, no statistical analysis was performed as the experiment was only done once.

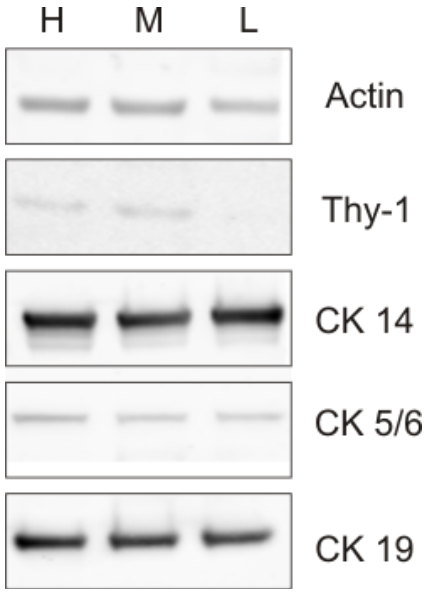


Figure 24 Western blot of Actin, Thy-1, CK14, CK5/6 and CK19

Actin was used as standard

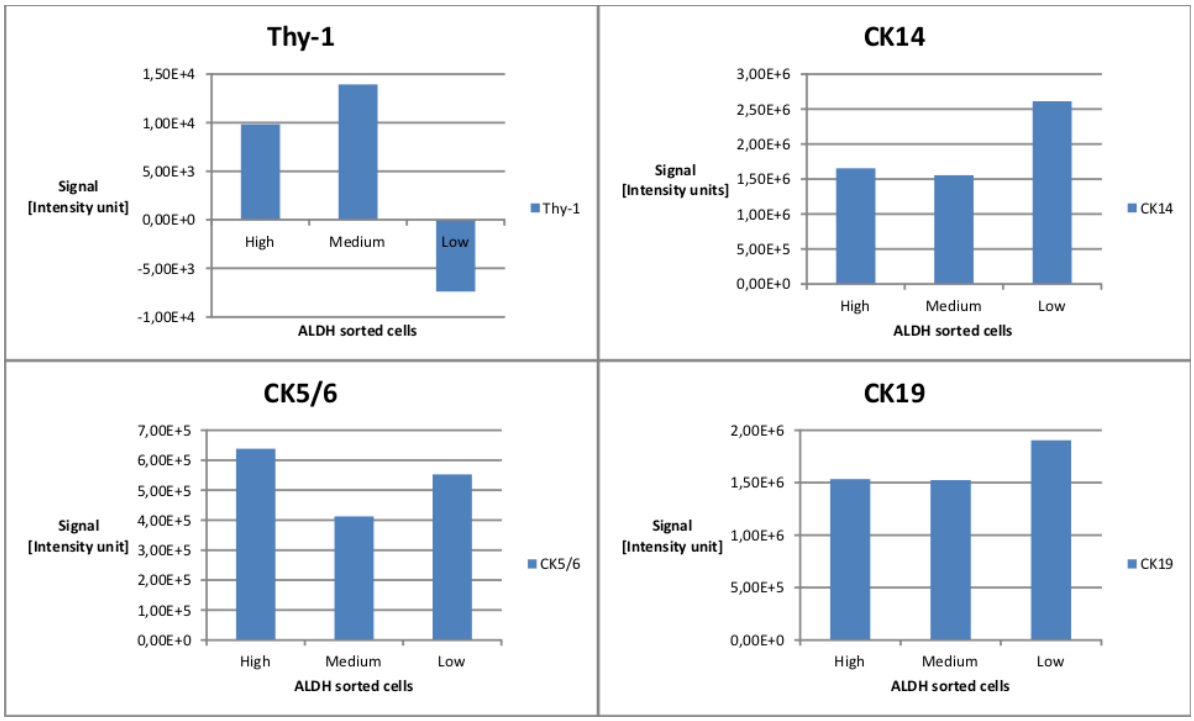


Figure 25 Western blot quantification data for expression of Thy-1, CK14, CK5/6 and CK19

4.6 ALDH high, medium and low cells grown in 3D culture

4.6.1 3D culture in Matrigel

ALDH sorted cell fractions were placed in 3D culture and co-cultured with and without BRENCs. On day 8 only a few small colonies had formed from ALDH high cells (Figure 26A) co-cultured with BRENCs in EGM5 medium. From ALDH medium and low cells with BRENCs only single viable cells or very small colonies were visible (Figure 26B and C).

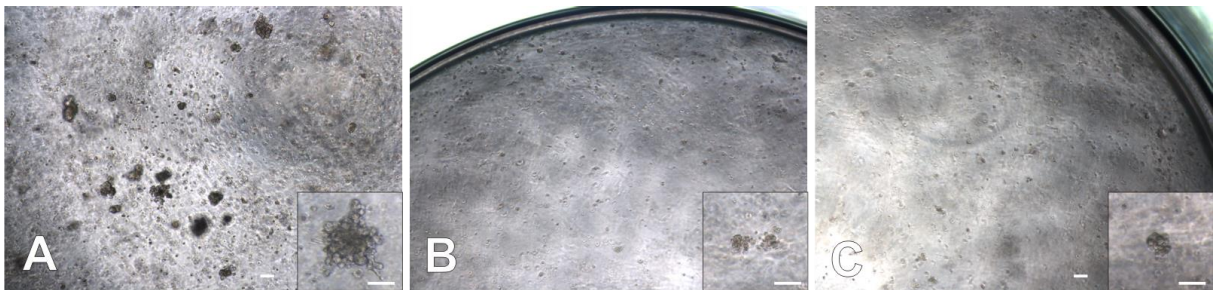


Figure 26 3D culture at day 8

A: ALDH high cells + BRENCs. **B:** ALDH medium cells + BRENCs. **C:** ALDH low cells + BRENCs. All cell types were grown in Matrigel with EGM5 medium. Phase contrast light microscopy, original magnification 50x and 100x in boxes.

Bar – 100 µm.

On day 16 a dramatic growth of colonies had occurred for all sorted cell types in co-culture with BRENCs. From ALDH high cells various types of colonies had formed; branching-like, budding, mesenchymal-like and solid (Figure 27). Most colonies from ALDH medium cells appeared smaller and only budding, not branching-like or mesenchymal- like (Figure 28). Colonies from ALDH low cells were similar in size as ALDH high colonies but did not branch as extensively (Figure 29). Clear cell aggregation was visualized in ALDH low (Figure 29C).

When sorted cell fractions were cultured without BRENCS, large and properly visible colonies were only seen in ALDH high cultures (Figure 30A). ALDH medium and low cells generated very small colonies or none at all (Figure 30B and C).

No colonies were formed on H14 medium regardless of sorted cell fraction.

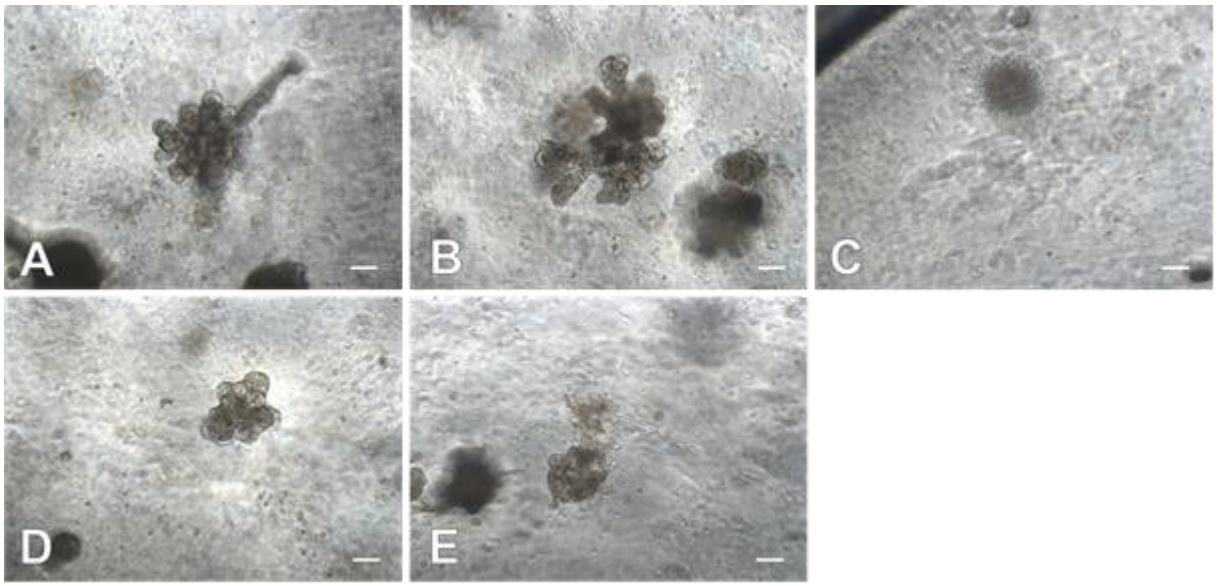


Figure 27 ALDH high cells with BRENCs in 3D culture after 16 days

A-B: Branching-like colonies. **C:** Mesenchymal-like colony. **D:** Budding colony. **E:** Solid and budding colony.

Colonies generated by ALDH high sorted cells with BRENCs under EGM5 medium. Phase contrast light microscopy, original magnification 100x. Bar – 100 µm.

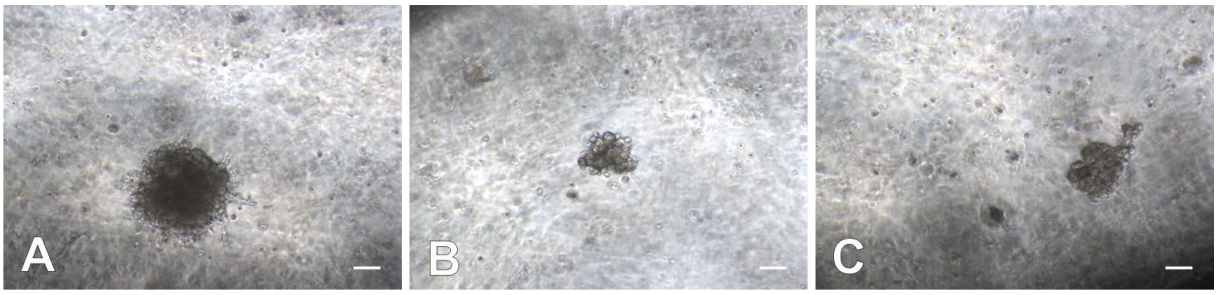


Figure 28 ALDH medium cells with BRENCs in 3D culture after 16 days

A-C: Various budding colonies. Colonies generated by ALDH medium sorted cells with BRENCs under EGM5 medium. Phase contrast light microscopy, original magnification 100x. Bar – 100 µm.

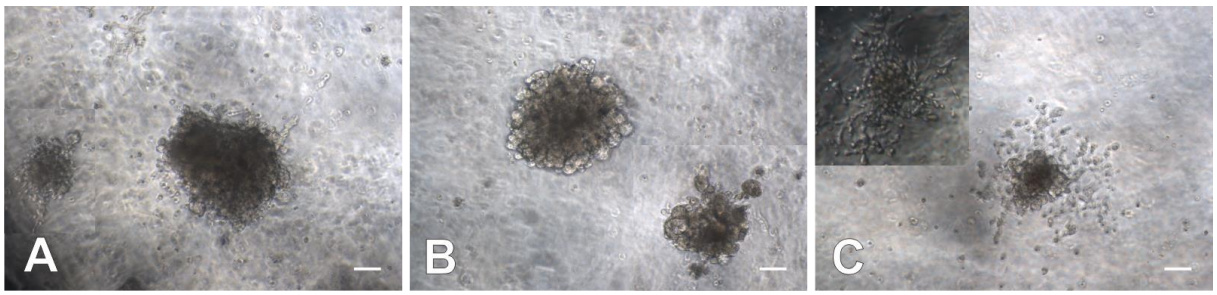


Figure 29 ALDH low cells with BRENCs in 3D culture after 16 days

A-B: Budding colonies. **C:** A mesenchymal-like colony (also inserted box).

Colonies generated by ALDH low sorted cells in co-culture with BRENCs under EGM5 medium. Phase contrast light microscopy, magnification 100x , Bar – 100 µm.

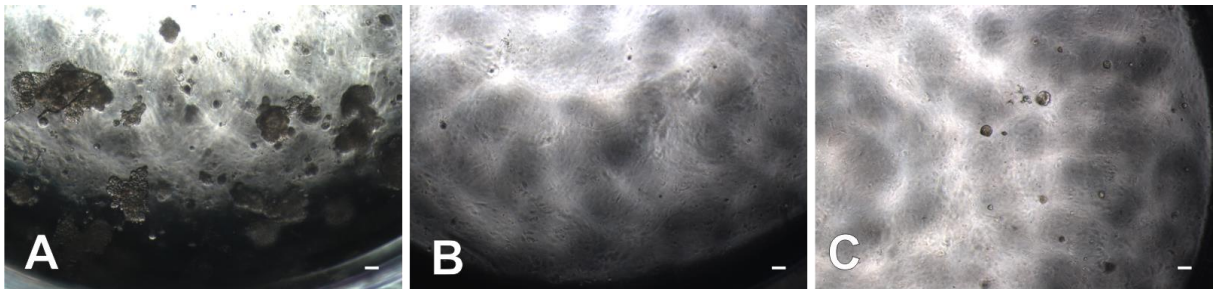


Figure 30 ALDH high, medium and low cells without BRENCs in 3D culture after 16 days

A: ALDH high cells. B: ALDH medium cells. C: ALDH low cells. Phase contrast light microscopy, original magnification 100x. Bar – 100 µm.

4.6.2 Immunofluorescence staining of colonies from 3D culture

Colonies isolated from 3D culture generated by sorted ALDH high, medium and low cells were stained for cytokeratins; 5/6, 8, 14, 17, 18 and 19. No results were obtained from staining of cytokeratin 5/6, 8, 17 and 18.

The colonies were additionally stained with F-actin for a better visualization of the structure (Figure 31 and 32). ALDH medium and low cells colonies stained with CK19 gave a stronger signal compared to ALDH high (Figure 31). In all cell types a strong signal from unspecific binding surrounding the colonies was observed.

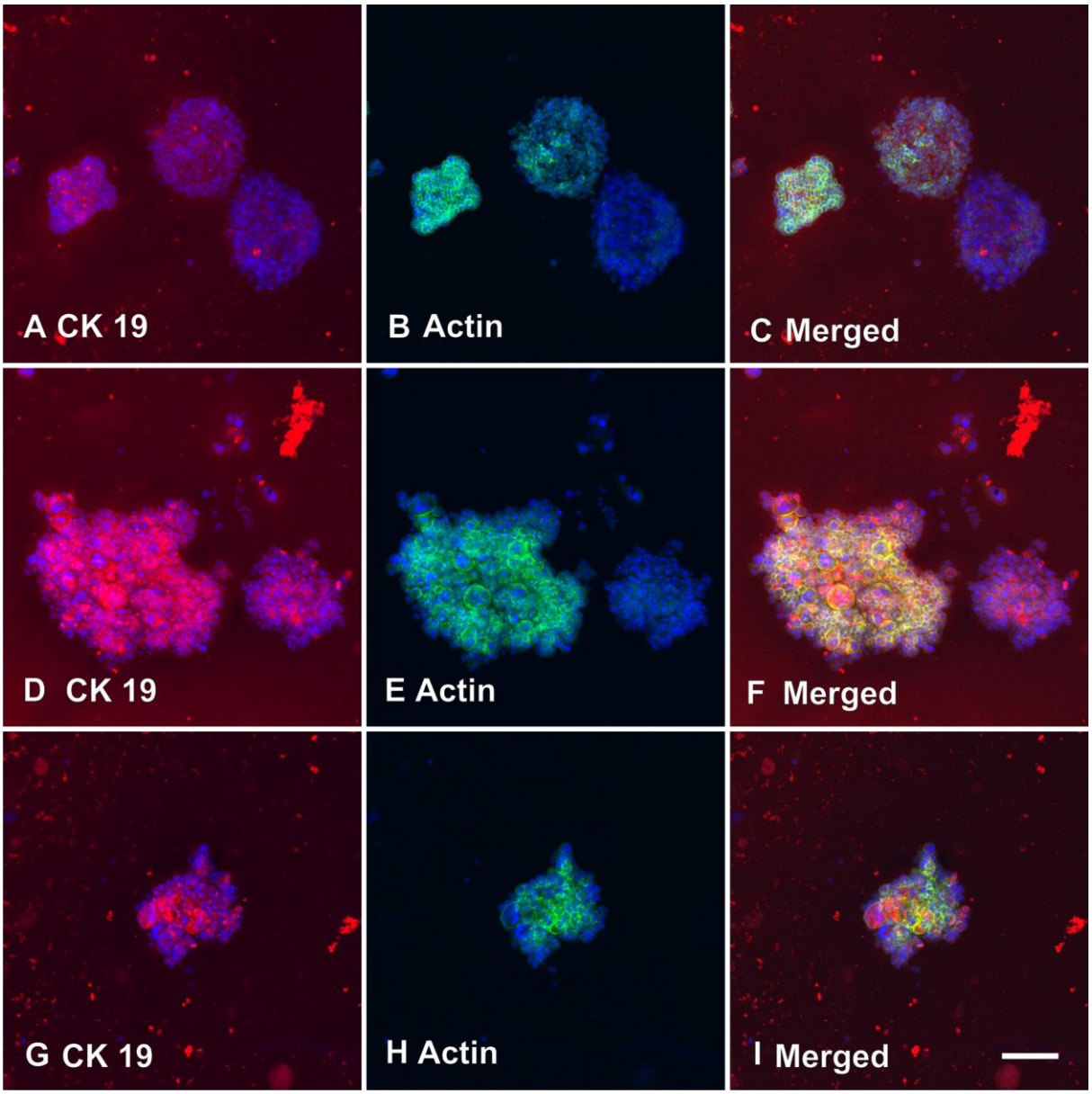


Figure 31 CK19 staining of colonies from ALDH high, medium and low cells with BRENCs

A-C: Colonies from ALDH sorted high cells. **D-F:** Colonies from ALDH sorted medium cells. **G-I:** A colony from ALDH sorted low cells. CK19 – red. F-actin – green. Nuclear counterstaining – blue (DAPI). Bar – 100 µm

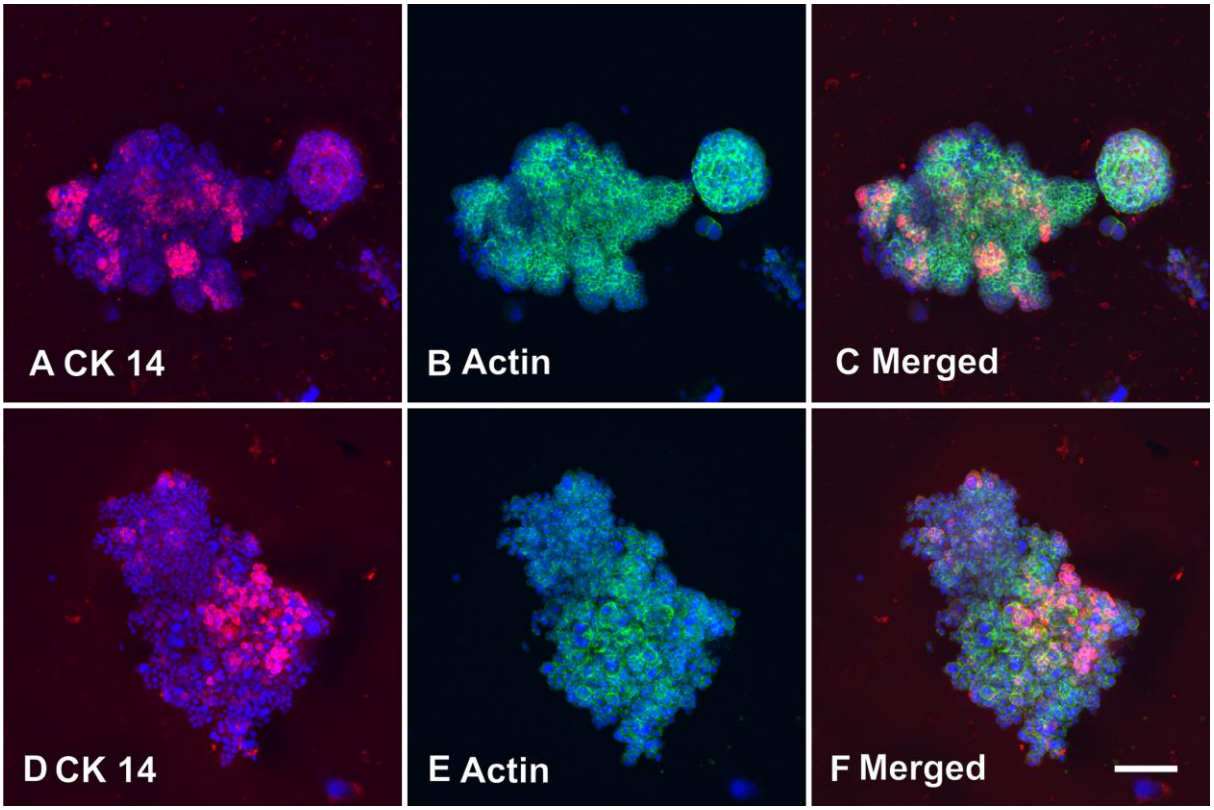


Figure 32 CK14 staining of colonies from ALDH high and low cells with BRENCs

A-C: Colonies from ALDH sorted high cells. **D-F:** A colony from ALDH sorted low cells. CK14 – red. F-actin – green. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.

CK14 gave only results for colonies from ALDH high and low cells; the staining was heterogeneous and similar between cell fractions (Figure 32).

Isolation and staining of colonies cultured from ALDH sorted cells without BRENCs gave only proper results in ALDH high cells; a branching-like colony and a mesenchymal-like colony (Figure 33A). The branching-like colony showed a strong and clear CK14 staining while the mesenchymal-like colony showed almost no CK14 staining. ALDH medium cells could only be evaluated by a single small colony positive for CK14 (Figure 33B). Isolation from ALDH low gave no colonies and only streaks of single cells were visualized, mostly CK14 positive. For both ALDH medium and low cells, high levels of unspecific antibody binding were observed (Figure 33B and C).

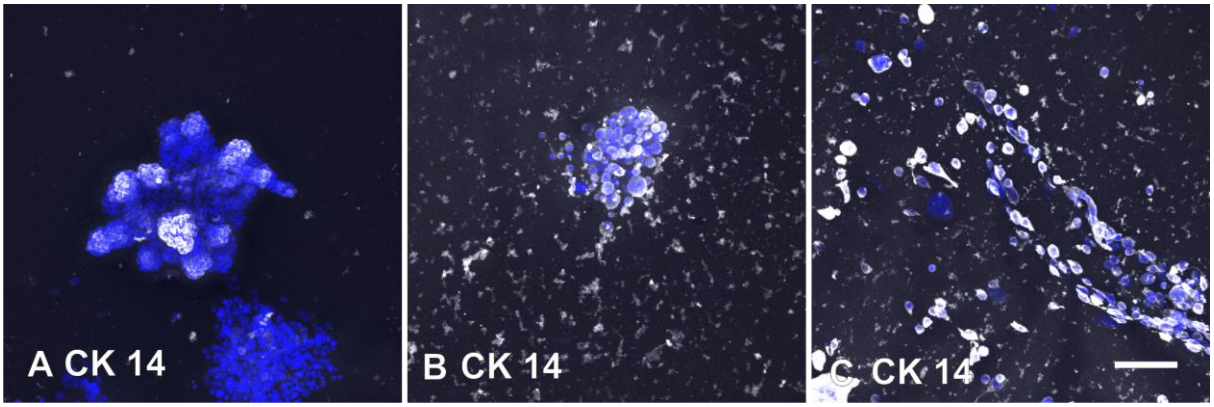


Figure 33 CK14 staining of colonies from ALDH high, medium and low cells without BRENCS

A: Colonies from ALDH sorted high cells. **B:** A colony from ALDH sorted medium cells. **C:** No colonies only cells isolated from ALDH sorted low cells. CK14 – white. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.

4.7 ALDH sorted fractions of D492 in monolayer

ALDH sorted cells grown in monolayer were stained for cytokeratins; 5/6, 8, 14, 17, 18 and 19. All sorted cell types showed heterogeneity in cytokeratin expression.

ALDH low cells showed a stronger CK14 staining than ALDH high and medium cells. CK18 and CK17 staining is similar in all cell types (Figure 34)

CK19 staining appeared stronger in ALDH high and medium cells compared to ALDH low cells while CK5/6 staining appeared similar (Figure 35)

CK8 staining was only performed in ALDH high and low cells (Figure 36). All ALDH low cells showed a staining of CK8 but the expression levels varied. Some ALDH high cells showed clearly no CK8 expression according to the staining.

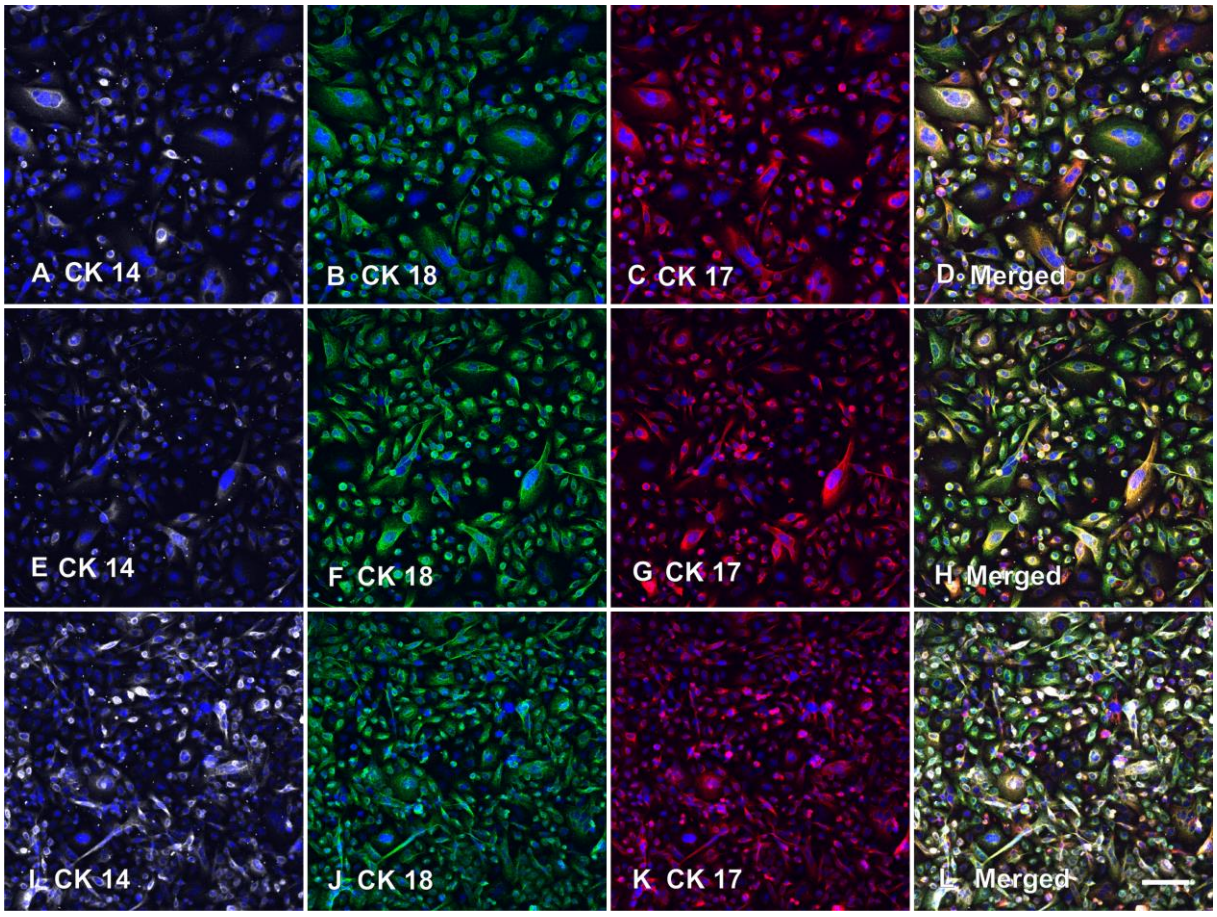


Figure 34 CK14, CK18 and CK17 staining of ALDH high, medium and low cells

A-D: ALDH sorted high cells. **E-H:** ALDH sorted medium cells. **I-L:** ALDH sorted low cells. CK14 – white. CK18 – green. CK17 – red. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.

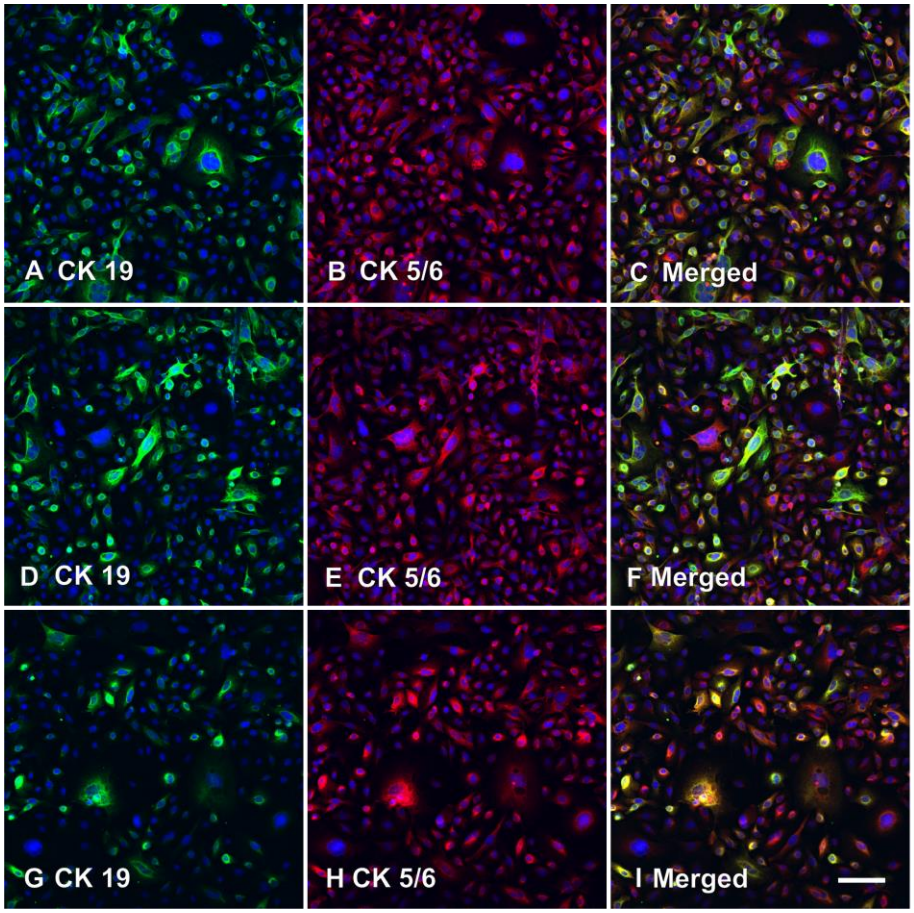


Figure 35 CK19 and CK5/6 staining of ALDH high, medium and low cells

A-C: ALDH sorted high cells. **D-F:** ALDH sorted medium cells. **G-I:** ALDH sorted low cells. CK19 – green. CK5/6 – red. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.

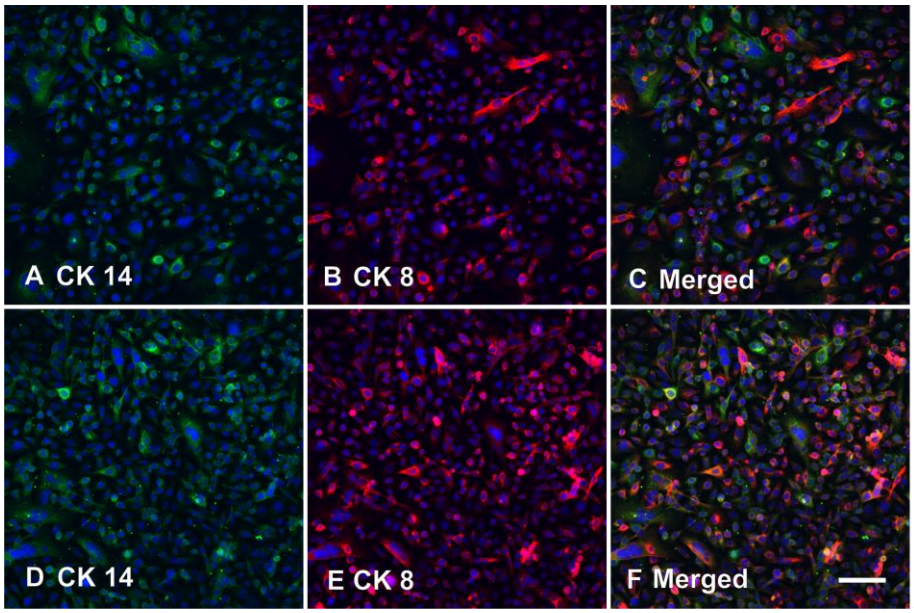


Figure 36 CK14 and CK8 staining of ALDH high and low cells

A-C: ALDH sorted high cells. **D-F:** ALDH sorted low cells. CK14 – green. CK8 – red. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.

5 Discussion

5.1 Summary

In this study ALDH activity was evaluated in relations to surface marker expression, cell passage number, protein expression and properties in 3D culture. ALDH activity was higher in D492 than the mesenchymal subline D492M. Generally surface marker expression was increased in ALDH high cell compared to ALDH low cells for both D492 and D492M. ALDH activity was not significantly affected by cell passage and ALDH sorted fractions showed minor differences in the expression of selected proteins. ALDH high cells proliferated faster than ALDH low cells when seeded at a low density. The results indicate that ALDH sorted fractions behave differently in 3D culture in relations to colony forming and branching, which could be attributed to increased stem cell properties in ALDH high cells. Further experiments are needed to confirm this.

5.2 Comparison of D492 and D492M

D492 showed significantly higher ALDH activity than D492M. D492M is a mesenchymal subline that has undergone EMT which is considered to increase stem cell properties (37, 38). It would therefore be reasonable to suggest increased ALDH activity in D492M. However others have demonstrated that ALDH activity is not increased in cells that have undergone EMT (39) supporting these findings.

In this experiment the comparison of D492 and D492M was not as direct as anticipated mostly due to instability and variable results of the negative control.

According to the manufacturer protocol for the Aldefluor assay the negative control should be used to set a gate on a dot plot with ALDH fluorescence against Side Scatter (SSC). This baseline gate is then applied to the test sample and fluorescence signals falling within represent cells with true high ALDH activity. When only one control and one test sample is measured this analysis gives a definite result. However when comparing ALDH activity between samples a problem arises since the negative controls give variable results. DEAB is the inhibitor of ALDH activity in negative controls and one would expect it to inhibit all activity and create a specific baseline which should be the same for all controls. This was however not the case for D492 when ALDH activity was compared between D492 and D492M.

The negative control was very unstable when analyzing D492 in triplicate and controls for D492M were also distinct. A decision was made to consistently compare ALDH activity between samples, by setting the gate for the negative control with the highest ALDH fluorescence signal. That gate was applied to all samples in order to obtain comparable results. This caused some samples to show a much lower ALDH activity than if the gate had been set according to their negative control. This however seemed to be the only way to compare the ALDH activity between different cultures, and by using a common gate results were at least consistent from sample to sample.

The reason for instability of the controls is difficult to determine. It is possible that some cellular metabolites (e.g. some aldehydes) bind to the ALDH and compete with the inhibitor and prevent it from binding to ALDH. Time could also be an influence as the method requires mixing the cell

suspension with Aldefluor Reagent and immediately transferring half of the volume to a tube containing DEAB. Perhaps during these seconds that vary slightly from control to control, uninhibited ALDH may convert BAAA in the Aldefluor Reagent to BAA, which has an effect on the negative control results. This is however merely speculations.

At this time point it is difficult to conclude how this problem should be addressed. Perhaps a more convenient way to evaluate ALDH would be to analyze gene expression of ALDH isoforms with Real time PCR between cell lines. However this could be problematic as it is not known which Aldehyde dehydrogenase isoforms contribute to the ALDH activity and experiments attempting to elucidate this are both time consuming and expensive.

An additional conflict in comparing D492 and D492M is SSC. D492 and D492M have differences in SSC properties i.a. D492M shows lower SSC than D492. This caused difficulties in setting live gates highlighting viable nucleated cells. For comparison of cell lines to be consistent identical gates were applied. This meant that in some cases outliers and other non-optimal cells were included in the live gate and others that technically should have been included were excluded. Obviously this was not optimal for direct comparison.

5.2.1 Marker expression and ALDH

The markers chosen for the study have been referred to as stem cell markers and used to enrich for stem cells in normal tissue but mostly CSCs. The idea was to map ALDH activity with respect to immunophenotyping observed with these markers; however this was not possible in most cases due to homogenous expression in D492. Surprisingly marker expression was quite similar between D492 and D492M. The markers selected can therefore not be used to characterize and enrich for stem cells in D492 or D492M. The only marker that discriminated between ALDH subpopulations was CD133.

$CD44^+$ and $CD24^-$ is a known immunophenotype in the literature concerning EMT and CSC. As D492M has undergone EMT it was expected that the CD24 expression would decrease and CD44 expression increase compared to D492. Decreased CD24 expression was described earlier by Sigurdsson et al. in D492M compared to D492, although in their analysis the CD44 expression appeared to decrease as well (37). In the contrary CD24 expression increased in D492M compared to D492 in this study. This inconsistency between experiments using the same cell lines is difficult to explain. The exact details behind the data of Sigurdsson et al. are unknown and multiple external factors could explain this difference.

A decrease in expression of E-Cadherin for cells that have undergone EMT is very well known and has been previously observed for D492M (37) and other cell types. The results in this experiment show a consistent diminished Cadherin expression in D492M as seen in Figure 10G. Thy-1 has also previously been shown to be increased in D492M which is supported by this data. However the difference between cell lines is perhaps not as great as expected since Thy-1 was previously demonstrated to be highly upregulated in D492M (37).

Comparison of marker expression in cells expressing 5% highest and lowest ALDH activity on the one hand and with the 5% highest and lowest SSC on the other, revealed similar effect on marker expression and almost exactly the same effect was seen for CD24 or CD117 against CD44

(Supplementary figures 8 and 9). Pluripotent stem cells have been shown to possess high SSC properties (63). It is possible that the same applies for more differentiated stem cells like MaSC. If confirmed that ALDH high is a surrogate for high SSC properties and vice versa while still reflecting stem cell character, the tasks of enriching for viable stem cells would be enormously simplified. Sorting on the basis of SSC would limit all effects on the cells caused by the Aldefluor assay or other stem cell assays. Although the substrate and product are non-toxic and the cells stay viable, the Aldefluor assay is likely to affect the cells to some extent.

The similarities of marker expression could also be related to the SSC. The high SSC, which basically reflects granularity of the cells may influence the fluorescence emitted by BAA (product generated by ALDH in the Aldefluor assay). The link between ALDH activity and SSC is in the very least worth looking into in future perspectives.

5.3 Cell passage and ALDH activity

Cell passage number seems not to dramatically influence ALDH activity in D492. The activity is lower in the later passages but not significantly compared to the earlier passages. If ALDH activity reflects stem cell character it may diminish with increased cell passage. This has been seen in other cell lines but albeit at much earlier passages (64). It is possible that at passage 30 and above the cells have reached certain stability in ALDH activity. Unfortunately earlier passages of D492 are difficult to obtain to demonstrate this. The D492 cells used in different studies mainly have a cell passage number in the range 30-60. It is therefore perhaps more relevant to know if the ALDH activity fluctuates in these passages.

5.4 Cell sorting in FACSAria

Sorting in FACSAria™ II was a hallmark for the experiment to analyze the difference between cells expressing ALDH high and low activity. However the cell sorter was unfortunately not optimized for breast epithelial cells such as D492.

This instrument is mostly used for sorting of bacteria and algae and is not in a routine use. Conditions were not sterile which is highly inappropriate as the sorted cells were harvested for 3D culture in rBM (Matrigel) and conventional monolayer culture.

Another uncertainty is the fact that there was a clear efflux of BAA (the fluorescence product) from the cells as the cells during sorting were at elevated temperatures. In order to maintain a selection of 10% ALDH high cells, the ALDH high gate had to be moved at regular time intervals. The effects of the efflux on further assays and results are undetermined but could be relevant. This may mean that the cells with high ALDH activity lost their fluorescence and were excluded from the sorting or sorted as ALDH medium or low cells.

To accelerate the sorting process as over 800.000 cells were sorted for each cell type the flow speed was kept quite high. This was possibly not favorable for the cells as they endured a high mechanical pressure during sorting. This may explain why only about 1/4 to 1/5 of the cells remained

viable after sorting. For the viable cells, the sorting caused no doubt tremendous cell stress which affected the cell morphology and very likely influenced metabolism resulting in higher ALDH activity.

5.4.1 ALDH activity after sorting

The number of cells seeded in monolayer culture clearly influenced maintenance of ALDH activity and ability to proliferate in ALDH sorted cells.

The data from ALDH activity measured in sorted cells after 10 days of culture indicates that when 5000 cells are seeded the original ALDH phenotype seems to be better maintained than for 1000 cells seeded. Seeding of 2000 cells displayed similar maintenance ability as 5000 seeded, but the difference between sorted cell fractions was not as distinguished.

The ALDH activity in 5000 cells increased when compared to how the activity was in unsorted cell that did not go through FACSAria sorting. This could indicate that when cells go through pressure of sorting the ALDH expression increases as a stress response. This conclusion can however not be made from that observation since not all cultures of 1000 and 2000 seeded cells showed increased expression compared to unsorted cells. In addition the ALDH activity in unsorted cells was measured straight after the Aldefluor assay while the sorted cells had been in culture for 10 days and therefore not as easily comparable.

Further evaluation of how sorting influences cellular stress and ALDH expression is an interesting future study. The cell morphology in monolayer after sorting was altered and characterized by large, multinucleated cells. This is most likely due to cellular stress resulting from mechanical pressure during the cell sorting. Interestingly even 10 days after sorting under stable culture conditions the cells appeared not to have fully recovered.

In order to evaluate proliferation ability, a simple cell counting was performed and compared to the original number of cells seeded. The greatest difference between sorted cell fractions was seen when 1000 cells were seeded with ALDH high cells expanding around 20 fold but ALDH low cells only 5-10 fold (manual and FACS counting distinguished). In a density as low as 1000 cells, close contact between cells is less frequent. The better proliferation ability of ALDH high cells could be an indicator of increased stem cells properties compared to ALDH low cells. With increased cell contact in cultures, cells seeded 2000 and especially 5000 showed less difference in proliferation between sorted cell fractions.

Cultures set up with 5000 seeded cells were the only performed in triplicate as cultures 1000 and 2000 cells seeded were used for protein isolation and monolayer staining. This restricts the possibility of statistical analysis and appropriate comparison between numbers of cells seeded.

5.4.2 Protein expression – Western blot

The most noticeable result of the western blot analysis is the absence of Thy-1 expression in ALDH low cells. Thy-1 is a known stem cell marker and this data therefore supports that stem cell properties accompany higher ALDH activity. However, additional experiments are needed to verify this

With increased stemness in ALDH high cells and connection to EMT, it is suggested that they might have the tendency to shift towards EMT or show a more EMT phenotype (34). Cells undergone

EMT show a decline in cytokeratin expression as previously described (37). Cytokeratin expression (CK14 and 19) is diminished in ALDH high cells but the difference is not definite enough to confirm this suggestion. E-Cad expression which decreases in EMT does not support this either as E-Cad is elevated in ALDH high cells. ALDH was also proposed as a marker of primitive luminal cells rather than stem/progenitor cells. The western blot results in this study cannot support that as the expression of luminal markers (CK19) is not elevated in ALDH high cells. However, the difference in E-Cad and CK19 expression between ALDH cell fractions is not decisive enough to fully reject these suggestions either.

The cells used for western blot were originally seeded at the density of 2000 cells and cultured for 10 days. Evaluation of ALDH activity in cultures revealed that these cells did not decisively maintain their ALDH phenotype as the 5000 cells seeded culture. The culture with 2000 cells seeded was thus not the most optimal culture to evaluate protein expression. However, that was not known beforehand when protein was isolated for western blot.

5.4.3 3D culture in Matrigel

The sorted ALDH cells were seeded to Matrigel with or without BRENCs. The cells were seeded to three wells for every sorted cell fraction, but the colonies consistently only formed in the third well. The plausible explanation is the cells were not sufficiently suspended in the Matrigel that is a viscous liquid when kept on ice. The cells were suspended in the Matrigel after a centrifugation step and possibly more cells may have stayed in the bottom until dispended in the third and last portion of the gel.

After 8 days in culture colonies had only formed from ALDH high cells and not ALDH medium and low cells. This could be an indicator of increased stem cell properties in ALDH high cells.

When the culture proceeded colonies started to form, from the other sorted cell fractions as well. On day 16, different types of colonies in variable sizes had formed from ALDH high, medium and low cells co-cultured with BRENCs.

The colonies from ALDH high cells appeared to branch more than colonies from other cell fraction which rather formed buds. A clear mesenchymal-like colony generated from ALDH high cells with BRENCs was seen in phase contrast microscope. A single mesenchymal-like colony was observed in ALDH low cells, however, its shape and appearance was not as clear as the one generated by ALDH high cells. BRENCs are known to induce formation of mesenchymal colonies in D492 (37). After isolation of colonies from the Matrigel a mesenchymal-like colony was also observed from ALDH high cells cultured without BRENCs. ALDH medium and low cells without BRENCs did not generate colonies while a branching-like colony was seen from ALDH high cells.

This preliminary data may suggest that ALDH high cells have a better ability to form branching and mesenchymal colonies with and without BRENCs that could be consequence of increased stem cell properties.

By staining the colonies for cytokeratins and F-actin, the heterogeneity in cytokeratin expression was elucidated and structure of the colonies further displayed. Visually colonies from ALDH high cells

appear to branch more than colonies from the other sorted cell fractions as was indicated by phase contrast microscope images.

Simultaneous CK14 staining of a branching-like and a mesenchymal-like colony side by side demonstrated diminished expression of CK14 in the mesenchymal structure as previously described by Sigurdsson et al. with D492M (37)

More images showing stained branching-like colonies from ALDH high cells can be seen in supplementary figures 22 and 23.

5.4.4 Cytokeratins in monolayer culture

Visually there appears not to be difference in cytokeratin expression between ALDH sorted high, medium and low cells in monolayer. Western blot is a better method to quantify cytokeratin expression, rather than by visually estimating expression with IF staining of cytokeratins. However, the staining of monolayer cultures reveals the heterogeneous pattern of the cytokeratin expression that is consistent in every sorted cell fraction.

6 Conclusion

In summary, this study suggests that the mammary stem cell line D492 can be sorted into functionally distinct cell populations on the basis of the ALDH properties. Cells expressing high ALDH displayed higher expression of the stem cell marker Thy-1. They formed a greater number of colonies with a branching tendency in 3D mono- and co-culture with BRENCs. ALDH high cells displayed higher proliferation rate when seeded at a low density. This can be interpreted as indications of increased stem cell properties in cells expressing high levels of ALDH. ALDH levels were higher in D492 than D492M and remained relatively consistent between different generations or passage numbers of the cell line. None of the known stem cell surface markers used in this study displayed heterogeneous expression that could be used to surrogate the ALDH high subpopulation in D492 and D492M.

This preliminary study provides a good base for further studies aimed at characterizing the cellular and molecular control of epithelial stem cell in the human breast.

6.1 Future perspectives

As stated before, this study needs to be followed up with additional methods in order to confirm the association between ALDH activity and stem cell properties in D492.

Firstly the cell sorting process needs to be repeated so the different fractions (high, medium and low) can be subjected to improved 3D culture conditions and mammosphere assay. This task will be easier in the coming months with a new cell sorter being implemented at the University Hospital. This means that sorting can be carried out under sterile conditions in an instrument optimized for sorting of breast epithelial cells. Some kind of a cooling system for the sample will be set up to avoid the product efflux. Due to the location, the transport time factor which likely affected cell survival after sorting will be eliminated.

The observed cellular stress due to mechanical pressure of sorting would be informative to study further. This may be possible to evaluate e.g. by monitoring reactive oxygen species (ROS) that are increased in conditions of cellular stress.

Analysis on gene expression to reveal which ALDH isoforms are expressed in D492 and D492M should be conducted, in order to identify the isoforms contributing to the fluorescence detected in the Aldefluor assay. Cells with high ALDH activity, if due to one or few major isoforms, could then be traced by the use of a reported gene construct carrying Green fluorescence protein (GFP). This setup would allow life imaging of these potential stem cells in culture and enable direct monitoring of their proliferation and behavior in 3D culture and response to certain chemicals and drugs.

Further analysis on the connection between ALDH activity and SSC should be perused. This can be tested by sorting cells on the basis of either ALDH activity or SSC and placing them in stem cells assays e.g. 3D culture and mammosphere assay. Since SSC can be monitored without chemical interference, it may become a useful tool to enrich for stem cells in D492 if it turns out to reflect ALDH activity and stemness.

The characterization of stem cell population in D492 and D492M with surface markers should be continued even though the markers in this study did not identify subpopulations with stem cell potentials.

Another challenge in this study is to evaluate ALDH activity in terms of chemotherapeutic drug resistance and in this respect monitor the response of D492 and D492M to chemotherapeutic drugs. The potential role of ALDH in drug resistance could be determined by inhibiting ALDH activity (65).

This preliminary study provides a very good base to build on for further analysis of ALDH activity in D492 and D492M in relations to stem cell properties, therapeutic drug resistance, cellular metabolism and stress response.

References

1. Sigurdsson V. Endothelial cells in breast morphogenesis. Reykjavik: University of Iceland; 2005.
2. Ingthorsson S. Modeling the role of the EGFR-receptor family in the normal and malignant breast gland. Reykjavik: University of Iceland; 2014.
3. Marcato P, Dean CA, Giacomantonio CA, Lee PW. Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform. *Cell Cycle*. 2011;10(9):1378-84.
4. Cai J, Weiss ML, Rao MS. In search of "stemness". *Exp Hematol*. 2004;32(7):585-98.
5. Burger PE, Gupta R, Xiong X, Ontiveros CS, Salm SN, Moscatelli D, et al. High aldehyde dehydrogenase activity: a novel functional marker of murine prostate stem/progenitor cells. *Stem Cells*. 2009;27(9):2220-8.
6. Ronnov-Jessen L, Petersen OW, Bissell MJ. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol Rev*. 1996;76(1):69-125.
7. Gudjonsson T, Villadsen R, Nielsen HL, Ronnov-Jessen L, Bissell MJ, Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev*. 2002;16(6):693-706.
8. Sigurdsson V. Cellular and molecular mechanism in breast morphogenesis and epithelial to mesenchymal transition Reykjavik: University of Iceland; 2012.
9. Kolios G, Moodley Y. Introduction to Stem Cells and Regenerative Medicine. *Respiration*. 2013;85(1):3-10.
10. Maehle AH. Ambiguous cells: the emergence of the stem cell concept in the nineteenth and twentieth centuries. *Notes Rec R Soc Lond*. 2011;65(4):359-78.
11. Appasani K, Appasani RK. Introduction to Stem Cells and Regenerative Medicine. In: Appasani K, Appasani RK, editors. Stem Cells and Regenerative Medicine. New York, NY: Humana Press; 2012.
12. Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. 1961. *Radiat Res*. 2012;178(2):Av3-7.
13. Burkert J, Wright NA, Alison MR. Stem cells and cancer: an intimate relationship. *J Pathol*. 2006;209(3):287-97.
14. Douville J, Beaulieu R, Balicki D. ALDH1 as a functional marker of cancer stem and progenitor cells. *Stem Cells Dev*. 2009;18(1):17-25.
15. Antonini S, Vescovi A, Gelain F. Stem Cells and Biomaterials: The Tissue Engineering Approach. In: Appasani K, Appasani RK, editors. Stem cells and Regenerative Medicine. New York, NY: Humana Press; 2011.
16. Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea--a paradigm shift. *Cancer Res*. 2006;66(4):1883-90; discussion 95-6.
17. Liu S, Dontu G, Wicha MS. Mammary stem cells, self-renewal pathways, and carcinogenesis. *Breast Cancer Res*. 2005;7(3):86-95.

18. Stingl J. Detection and analysis of mammary gland stem cells. *J Pathol*. 2009;217(2):229-41.

19. Pechoux C, Gudjonsson T, Ronnov-Jessen L, Bissell MJ, Petersen OW. Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. *Dev Biol*. 1999;206(1):88-99.

20. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, et al. Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol*. 2007;177(1):87-101.

21. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, et al. Generation of a functional mammary gland from a single stem cell. *Nature*. 2006;439(7072):84-8.

22. Virchow R. Editorial. *Virchows Arch Pathol Anat Physiol Klin Med*. 1855;3:23.

23. Roy S, Majumdar AP. Signaling in colon cancer stem cells. *J Mol Signal*. 2012;7(1):11.

24. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A*. 2003;100(7):3983-8.

25. Kakarala M, Wicha MS. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. *J Clin Oncol*. 2008;26(17):2813-20.

26. Yu Z, Pestell TG, Lisanti MP, Pestell RG. Cancer stem cells. *Int J Biochem Cell Biol*. 2012;44(12):2144-51.

27. Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, et al. Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res*. 2006;66(19):9339-44.

28. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell*. 2007;1(5):555-67.

29. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. 1997;3(7):730-7.

30. Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. *Cancer Res*. 2003;63(18):5821-8.

31. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J Natl Cancer Inst*. 2008;100(9):672-9.

32. McCubrey JA, Abrams SL, Fitzgerald TL, Cocco L, Martelli AM, Montalto G, et al. Roles of signaling pathways in drug resistance, cancer initiating cells and cancer progression and metastasis. *Adv Biol Regul*. 2014.

33. Kotiyal S, Bhattacharya S. Breast cancer stem cells, EMT and therapeutic targets. *Biochem Biophys Res Commun*. 2014.

34. May CD, Sphyris N, Evans KW, Werden SJ, Guo W, Mani SA. Epithelial-mesenchymal transition and cancer stem cells: a dangerously dynamic duo in breast cancer progression. *Breast Cancer Res*. 2011;13(1):202.

35. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell*. 2009;139(5):871-90.

36. Foroni C, Broggini M, Generali D, Damia G. Epithelial-mesenchymal transition and breast cancer: role, molecular mechanisms and clinical impact. *Cancer Treat Rev*. 2012;38(6):689-97.

37. Sigurdsson V, Hilmarsdottir B, Sigmundsdottir H, Fridriksdottir AJ, Ringner M, Villadsen R, et al. Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One*. 2011;6(9):e23833.

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

39. Sarrio D, Franklin CK, Mackay A, Reis-Filho JS, Isacke CM. Epithelial and mesenchymal subpopulations within normal basal breast cell lines exhibit distinct stem cell/progenitor properties. *Stem Cells*. 2012;30(2):292-303.

40. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts. *Stem Cell Reports*. 2014;2(1):78-91.

41. Vasiliou V, Thompson DC, Smith C, Fujita M, Chen Y. Aldehyde dehydrogenases: from eye crystallins to metabolic disease and cancer stem cells. *Chem Biol Interact*. 2013;202(1-3):2-10.

42. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol*. 2008;4(6):697-720.

43. Vasiliou V, Pappa A, Petersen DR. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem Biol Interact*. 2000;129(1-2):1-19.

44. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. *Stem Cells*. 2011;29(1):32-45.

45. Graham CE, Brocklehurst K, Pickersgill RW, Warren MJ. Characterization of retinaldehyde dehydrogenase 3. *Biochem J*. 2006;394(Pt 1):67-75.

46. Chute JP, Muramoto GG, Whitesides J, Colvin M, Safi R, Chao NJ, et al. Inhibition of aldehyde dehydrogenase and retinoid signaling induces the expansion of human hematopoietic stem cells. *Proc Natl Acad Sci U S A*. 2006;103(31):11707-12.

47. Fan X, Molotkov A, Manabe S, Donmoyer CM, Deltour L, Foglio MH, et al. Targeted disruption of Aldh1a1 (Raldh1) provides evidence for a complex mechanism of retinoic acid synthesis in the developing retina. *Mol Cell Biol*. 2003;23(13):4637-48.

48. Sladek NE. Aldehyde dehydrogenase-mediated cellular relative insensitivity to the oxazaphosphorines. *Curr Pharm Des*. 1999;5(8):607-25.

49. Charafe-Jauffret E, Ginestier C, Iovino F, Tarpin C, Diebel M, Esterni B, et al. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin Cancer Res*. 2010;16(1):45-55.

50. Dupe V, Matt N, Garnier JM, Chambon P, Mark M, Ghyselinck NB. A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. *Proc Natl Acad Sci U S A*. 2003;100(24):14036-41.

51. Rexer BN, Zheng WL, Ong DE. Retinoic acid biosynthesis by normal human breast epithelium is via aldehyde dehydrogenase 6, absent in MCF-7 cells. *Cancer Res*. 2001;61(19):7065-70.

52. Gentry T, Foster S, Winstead L, Deibert E, Fiordalisi M, Balber A. Simultaneous isolation of human BM hematopoietic, endothelial and mesenchymal progenitor cells by flow sorting based on aldehyde dehydrogenase activity: implications for cell therapy. *Cytotherapy*. 2007;9(3):259-74.

53. Cheung AM, Wan TS, Leung JC, Chan LY, Huang H, Kwong YL, et al. Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential. *Leukemia*. 2007;21(7):1423-30.

54. Eirew P, Kannan N, Knapp DJ, Vaillant F, Emerman JT, Lindeman GJ, et al. Aldehyde dehydrogenase activity is a biomarker of primitive normal human mammary luminal cells. *Stem Cells*. 2012;30(2):344-8.

55. Storms RW, Trujillo AP, Springer JB, Shah L, Colvin OM, Ludeman SM, et al. Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proc Natl Acad Sci U S A*. 1999;96(16):9118-23.

56. Levi BP, Yilmaz OH, Duester G, Morrison SJ. Aldehyde dehydrogenase 1a1 is dispensable for stem cell function in the mouse hematopoietic and nervous systems. *Blood*. 2009;113(8):1670-80.

57. Li ML, Aggeler J, Farson DA, Hatier C, Hassell J, Bissell MJ. Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. *Proc Natl Acad Sci U S A*. 1987;84(1):136-40.

58. Kleinman HK, Martin GR. Matrigel: basement membrane matrix with biological activity. *Semin Cancer Biol*. 2005;15(5):378-86.

59. Shaw FL, Harrison H, Spence K, Ablett MP, Simoes BM, Farnie G, et al. A detailed mammosphere assay protocol for the quantification of breast stem cell activity. *J Mammary Gland Biol Neoplasia*. 2012;17(2):111-7.

60. Manuel Iglesias J, Beloqui I, Garcia-Garcia F, Leis O, Vazquez-Martin A, Eguiara A, et al. Mammosphere formation in breast carcinoma cell lines depends upon expression of E-cadherin. *PLoS One*. 2013;8(10):e77281.

61. Aldefluor™: Aldehyde Dehydrogenase - Based Cell Detection kit, Product information sheet STEMCELL Technologies; 2011.

62. Lee GY, Kenny PA, Lee EH, Bissell MJ. Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat Methods*. 2007;4(4):359-65.

63. Ramirez JM, Bai Q, Pequignot M, Becker F, Kassambara A, Bouin A, et al. Side scatter intensity is highly heterogeneous in undifferentiated pluripotent stem cells and predicts clonogenic self-renewal. *Stem Cells Dev*. 2013;22(12):1851-60.

64. Cortes-Dericks L, Froment L, Boesch R, Schmid RA, Karoubi G. Cisplatin-resistant cells in malignant pleural mesothelioma cell lines show ALDH(high)CD44(+) phenotype and sphere-forming capacity. *BMC Cancer*. 2014;14:304.

65. Raha D, Wilson TR, Peng J, Peterson D, Yue P, Evangelista M, et al. The cancer stem cell marker aldehyde dehydrogenase is required to maintain a drug-tolerant tumor cell subpopulation. *Cancer Res*. 2014;74(13):3579-90.

66. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*. 2009;15(8):907-13.

67. Donnenberg VS, Donnenberg AD, Zimmerlin L, Landreneau RJ, Bhargava R, Wetzel RA, et al. Localization of CD44 and CD90 positive cells to the invasive front of breast tumors. *Cytometry B Clin Cytom*. 2010;78(5):287-301.

68. Kadmon G, Eckert M, Sammar M, Schachner M, Altevogt P. Nectadrin, the heat-stable antigen, is a cell adhesion molecule. *J Cell Biol*. 1992;118(5):1245-58.

69. Hubbe M, Altevogt P. Heat-stable antigen/CD24 on mouse T lymphocytes: evidence for a costimulatory function. *Eur J Immunol*. 1994;24(3):731-7.

70. Lingala S, Cui YY, Chen X, Ruebner BH, Qian XF, Zern MA, et al. Immunohistochemical staining of cancer stem cell markers in hepatocellular carcinoma. *Exp Mol Pathol*. 2010;89(1):27-35.

71. Zhu H, Mitsuhashi N, Klein A, Barsky LW, Weinberg K, Barr ML, et al. The role of the hyaluronan receptor CD44 in mesenchymal stem cell migration in the extracellular matrix. *Stem Cells*. 2006;24(4):928-35.

72. Liu AY, True LD, LaTray L, Nelson PS, Ellis WJ, Vessella RL, et al. Cell-cell interaction in prostate gene regulation and cytodifferentiation. *Proc Natl Acad Sci U S A*. 1997;94(20):10705-10.

73. Li Z. CD133: a stem cell biomarker and beyond. *Exp Hematol Oncol*. 2013;2(1):17.

74. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, et al. Identification of human brain tumour initiating cells. *Nature*. 2004;432(7015):396-401.

75. Castillo V, Valenzuela R, Huidobro C, Contreras HR, Castellon EA. Functional characteristics of cancer stem cells and their role in drug resistance of prostate cancer. *Int J Oncol*. 2014;45(3):985-94.

76. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*. 2007;445(7123):106-10.

77. Kurata R, Futaki S, Nakano I, Tanemura A, Murota H, Katayama I, et al. Isolation and characterization of sweat gland myoepithelial cells from human skin. *Cell Struct Funct*. 2014;39(2):101-12.

78. Rege TA, Hagood JS. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. *Faseb j*. 2006;20(8):1045-54.

79. Rege TA, Hagood JS. Thy-1, a versatile modulator of signaling affecting cellular adhesion, proliferation, survival, and cytokine/growth factor responses. *Biochim Biophys Acta*. 2006;1763(10):991-9.

80. Craig W, Kay R, Cutler RL, Lansdorp PM. Expression of Thy-1 on human hematopoietic progenitor cells. *J Exp Med*. 1993;177(5):1331-42.

81. Lobba AR, Forni MF, Carreira AC, Sogayar MC. Differential expression of CD90 and CD14 stem cell markers in malignant breast cancer cell lines. *Cytometry A*. 2012;81(12):1084-91.

82. Litvinov SV, Velders MP, Bakker HA, Fleuren GJ, Warnaar SO. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J Cell Biol*. 1994;125(2):437-46.

83. Stingl J, Eaves CJ, Zandieh I, Emerman JT. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat*. 2001;67(2):93-109.

84. Nakajima S, Doi R, Toyoda E, Tsuji S, Wada M, Koizumi M, et al. N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma. *Clin Cancer Res*. 2004;10(12 Pt 1):4125-33.

85. Truong Quang BA, Mani M, Markova O, Lecuit T, Lenne PF. Principles of E-cadherin supramolecular organization in vivo. *Curr Biol*. 2013;23(22):2197-207.

86. Hazan RB, Phillips GR, Qiao RF, Norton L, Aaronson SA. Exogenous expression of N-cadherin in breast cancer cells induces cell migration, invasion, and metastasis. *J Cell Biol*. 2000;148(4):779-90.

87. Franke WW, Schmid E, Osborn M, Weber K. Intermediate-sized filaments of human endothelial cells. *J Cell Biol*. 1979;81(3):570-80.

7 Appendix

7.1 Markers for characterazation of D492 and D492M

7.1.1 Stem cell surface markers

7.1.1.1 CD117

CD117 also known as c-kit, is a marker for progenitors of luminal breast cells but not adult cells. High expression of CD117 correlates with BRCA1 mutation (66). CD117 expression is also common in myeloid leukemia and gastrointestinal tumors. CD117 is a cytokine receptor, functions as a signaling molecule that affects survival, proliferation and differentiation of hematopoietic stem cells among other cell types (67).

7.1.1.2 CD24 and CD44

CD24 or heat-stable antigen (HAS) is a membrane bound glycoprotein with a molecular weight of 38-70 kDa. It mediates cell to cell adhesion presumably by self-binding (68).

CD24 can be used to identify neural stem cells in mice ($CD24^{lo/-}$) and is expressed on hematopoietic cells (21, 69).

CD44 is a cell surface adhesion molecule and a hyaluronic acid (HA) receptor with the predicted molecular weight of 82 kDa. CD44 is a marker for CSC in leukemia, breast, prostate, pancreatic and head or neck cancer (70). In normal tissue CD44 is expressed on mesenchymal stem cells and facilitates their migration when bound to HA. It may also affect homing (cells ability to migrate to target organs) (71). Basal cells in the prostate that have been shown to possess stem cell characteristics and being progenitors of luminal epithelial cells, express CD44 (72).

$CD24^-$, $CD44^+$ and Lin^- marker were also used by Al-Hajj et al. to isolate tumor cells from human breast tumors. These cells were able to give rise to tumorigenic cells with the same expression and also non-tumorigenic cells which were phenotypically different and made up the bulk of the tumor. These findings demonstrate CD44 and CD24 as CSC markers (24).

$CD44^+$ and $CD24^-$ have also been used to isolate mesenchymal-appearing cells from normal breast epithelium cells that have undergone EMT. These cells were identified as bipotential precursor cells, generating both luminal epithelial cells (expressed CK18) and basal/myoepithelial cells (expressed CK14) with a small subpopulation expressing both luminal and basal/myopepithelial markers (CK18 and CK14). These results pinpointed the $CD44^+/CD24^-$ population as stem/precursor cells in normal breast epithelium (38).

7.1.1.3 CD133

CD133 or Prominin-1 is a cell surface glycoprotein with five transmembrane domains (70). It has the molecular weight of 120 kDa. CD133 was originally discovered on hematopoietic stem cells (human) and neural epithelial cells (murine). Subsequently, expression on stem and/or progenitor cells was demonstrated and associated with differentiation, regeneration, metabolism and cancer phenotype marking it as a key marker for stem cells isolation and characterization (73). CD133 is a commonly

used marker in identification of CSC in various organs including brain (74), prostate (75) and colon (76).

7.1.1.4 CD29

CD29 or β1- intergrin was used by Shackelton et al. to isolate murine mammary stem cells (MaSC) cells. The subpopulation expressing CD29 along with $CD24^+$ and Lin^- was highly enriched for mammary stem cells defined by their self-renewal and multipotent properties. This was demonstrated by cell transplantation in mice where these cells were able to reconstruct a functional mammary gland and give rise to both luminal and myoepithelial lineages (21). CD29 has also been suggested as an epidermis stem cell marker. When expressed along with CD49f it defines a subpopulation considered to be sweat gland stem cells which are multipotent with the ability to form 3D spheres similar to mammospheres (77).

7.1.1.5 CD90 (Thy-1)

CD90, more commonly known as Thy-1, is a glycosylphosphatidylinositol-anchored glycoprotein with the molecular weight of 22-37 kDa. It is associated with T-cell activation, tumor suppression, apoptotic signaling, adhesion, migration, fibrosis and wound healing. Thy-1 is expressed on different cell types such as fibroblasts, endothelial and hematopoietic cells including stem cells, neurons and ovarian cancer cells (78, 79). Thy-1 is co-expressed with CD34 on human and murine hematopoietic primitive cells. As these stem/progenitor cells differentiate and commit to various lineages the Thy-1 expression decreases. This suggests that Thy-1 is a key lineage-independent marker for stem cells in adult tissue (67, 80).

Lobba et.al identified Thy-1 as a potential breast CSC marker following characterization of a malignant cell line (Hs578-T, isolated from primary breast tissue) with high expression of Thy-1. This cell line displays a high tumorigenic and metastatic potential. Thy-1 has been used as a marker for mesenchymal stem cells with its linkage to development, cancer and CSC (81).

In normal breast tissue Donnenberg et al. revealed that Thy-1 co-expression with CD44 is localized to basal ductal cells, but Thy-1 expression is lost in luminal cells and CD44 decreased. In breast tumors, Thy-1 cells with co-expression of $cytokeratin^+$ (pan-cytokeratins – CK5/6 and 8) and $CD44^+$ were localized to the periphery of tumors. They hypothesize that these cells are equivalent to the basal cells and suggest that they are the candidate breast cancer stem cells (67).

7.1.1.6 Epithelial cell adhesion molecule (EpCAM)

EpCAM, also known as epithelial-specific antigen (ESA) is an adhesion molecule and transmembrane glycoprotein with molecular weight around 40 kDa. EpCAM is expressed on the apical surface of luminal epithelial cells including those of the mammary gland (7, 82). EpCAM plays a role in cellular adhesion and forms tight junctions by regulating claudin proteins. The function of EpCAM is also associated with proliferation. Gudjonsson et al. and Stingl et al. found EpCAM expression on a subpopulation of luminal epithelial cells that could generate luminal and myoepithelial cells, suggesting a stem cell/progenitor phenotype (7, 83).

7.1.2 Other surface markers

7.1.2.1 Cadherins

E-Cadherin (E-Cad, CD324) and N-Cadherin (N-Cad, CD325) are epithelial surface markers and calcium-dependent adhesion molecules. They are involved in embryonic development and maintenance of tissue architecture (84).

E-Cad plays a role in supporting both formation and stabilization of cell contacts by interacting with the cytoskeleton. E-Cad prevents disruption of cell contacts when cells are under physical stress during tissue remodeling (85). In EMT, cells display a loss or reduction in E-Cad expression, linking its function to prevention of metastasis and tumor progression while N-Cad expression is increased (37, 84).

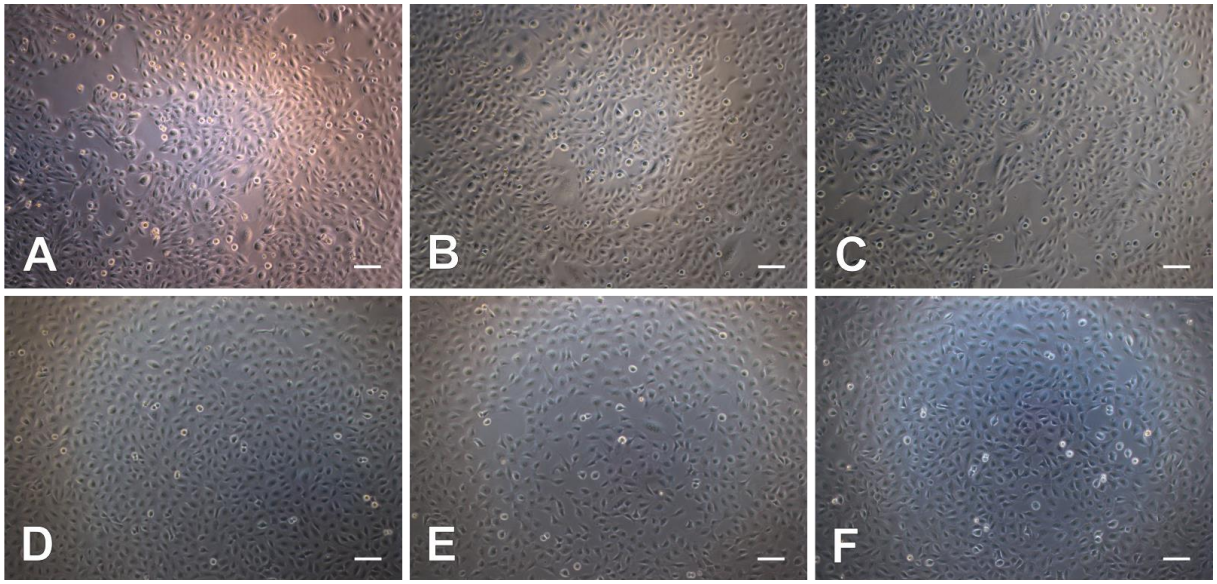
N-Cad is associated with increased invasiveness in cancer. N-Cad promotes motility and presumably has a more important role than E-Cad when it comes to invasiveness and metastasis. This was demonstrated by overexpressing N-Cad in usually E-Cad expressing, weak invasive and metastatic cell line. The cell line showed increased invasiveness and metastatic properties despite continually expressing E-Cad (86).

7.1.3 Intracellular markers

7.1.3.1 Cytokeratins

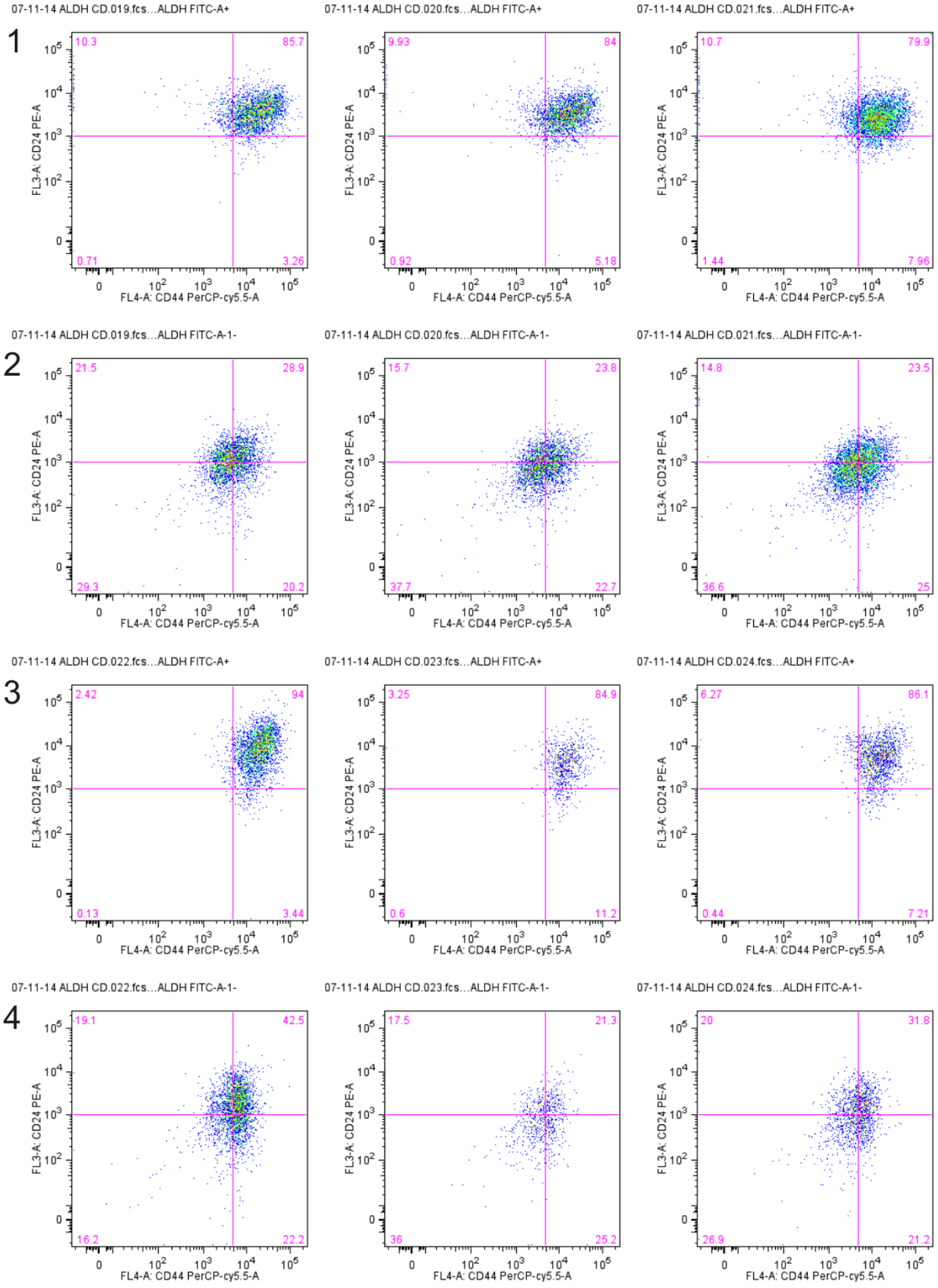
Cytokeratins (CK) are intermediate-sized filaments contributing to the cytoskeleton of epithelial cells (87). In breast tissue, luminal epithelial cells express CK8, CK18 and CK19 while myoepithelial cells express CK5, CK14 and CK17 (8).

7.2 Supplementary figures



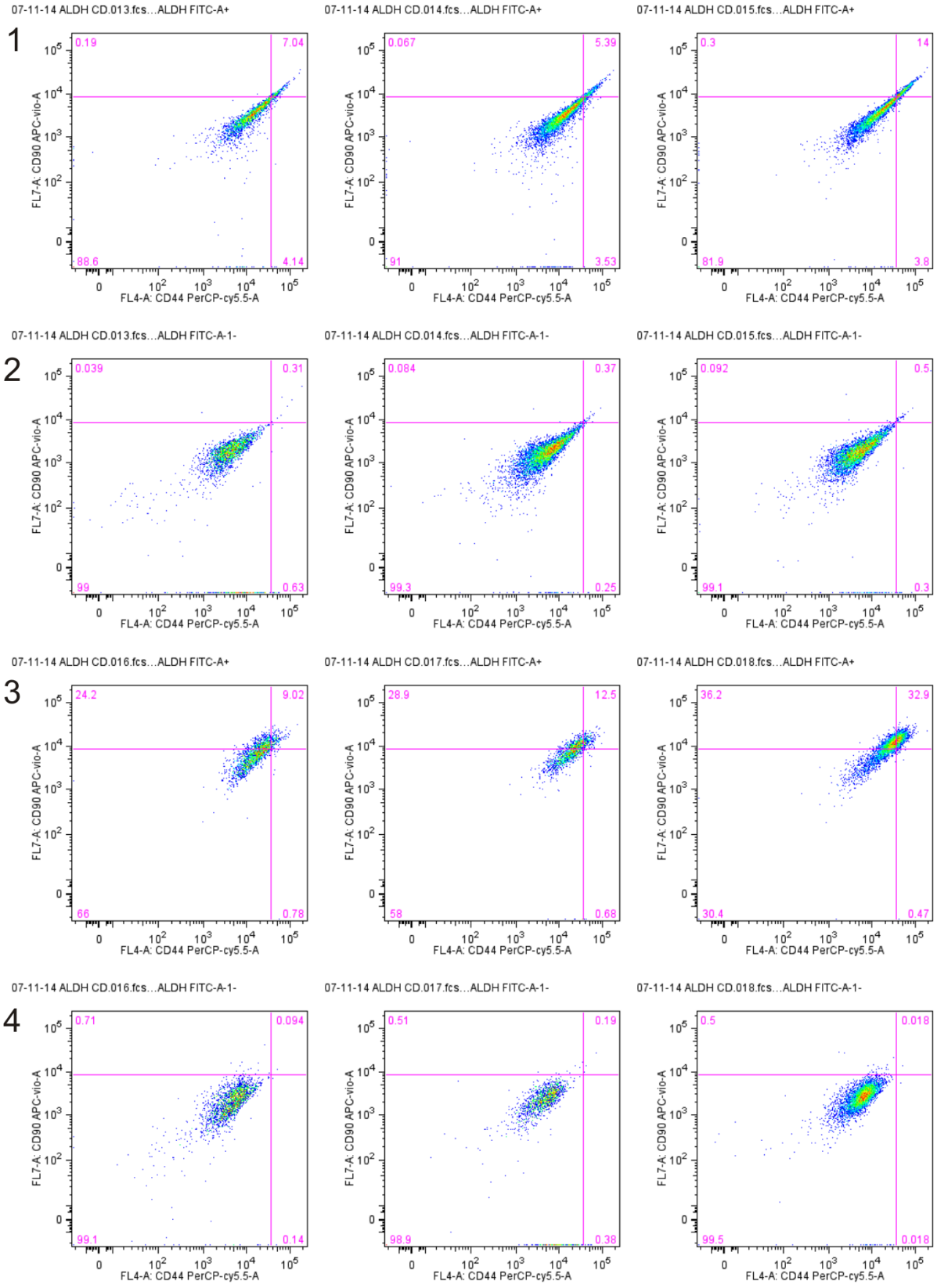
Supplementary figure 1 D492 and D492M in monolayer culture

A-C: D492 passage 36, wells 1-3. **D-F:** 492M passage 25, wells 1-3. Phase contrast light microscopy, original magnification 50x. Bar – 100 µm.



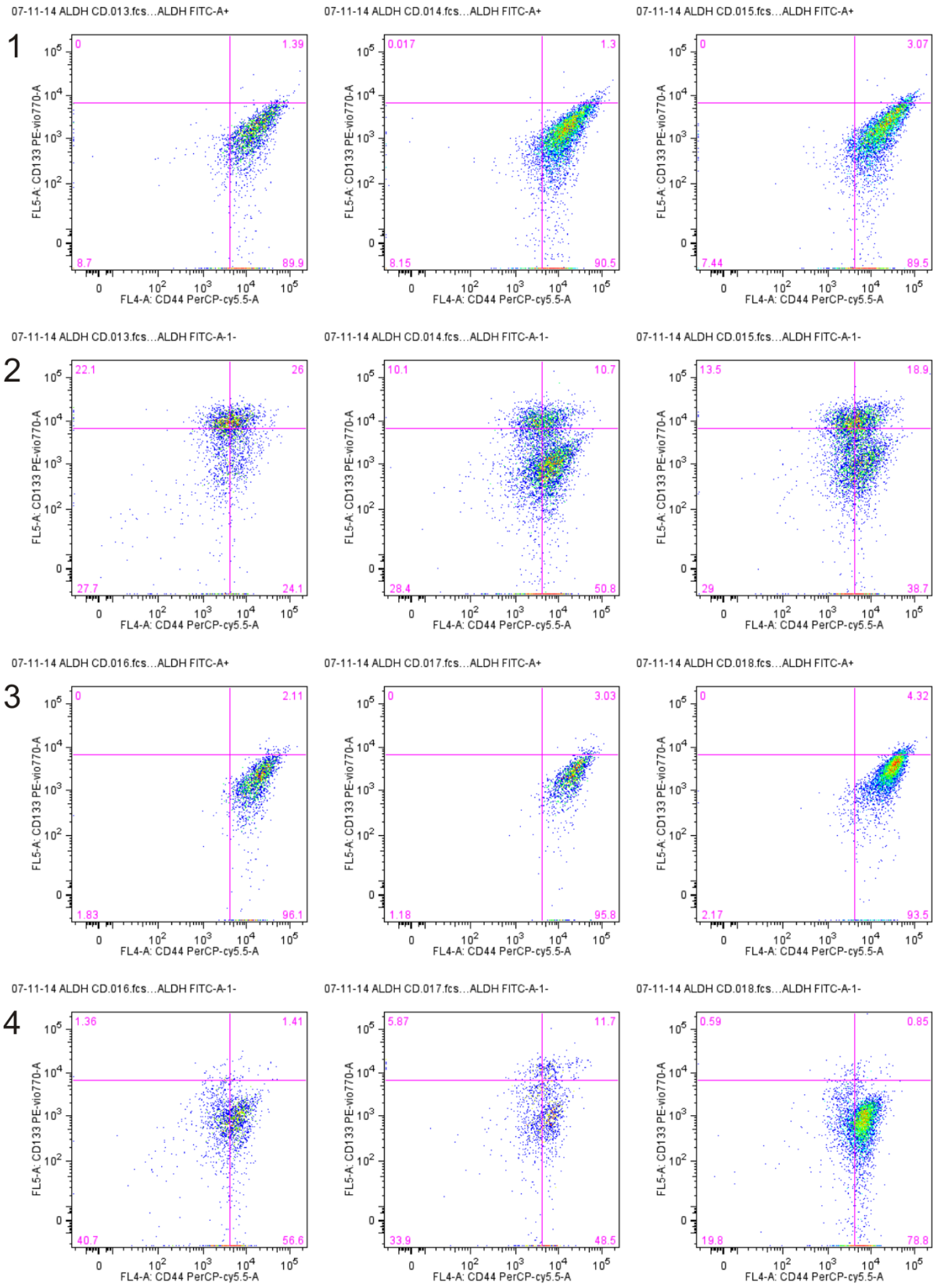
Supplementary figure 2 CD24 expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



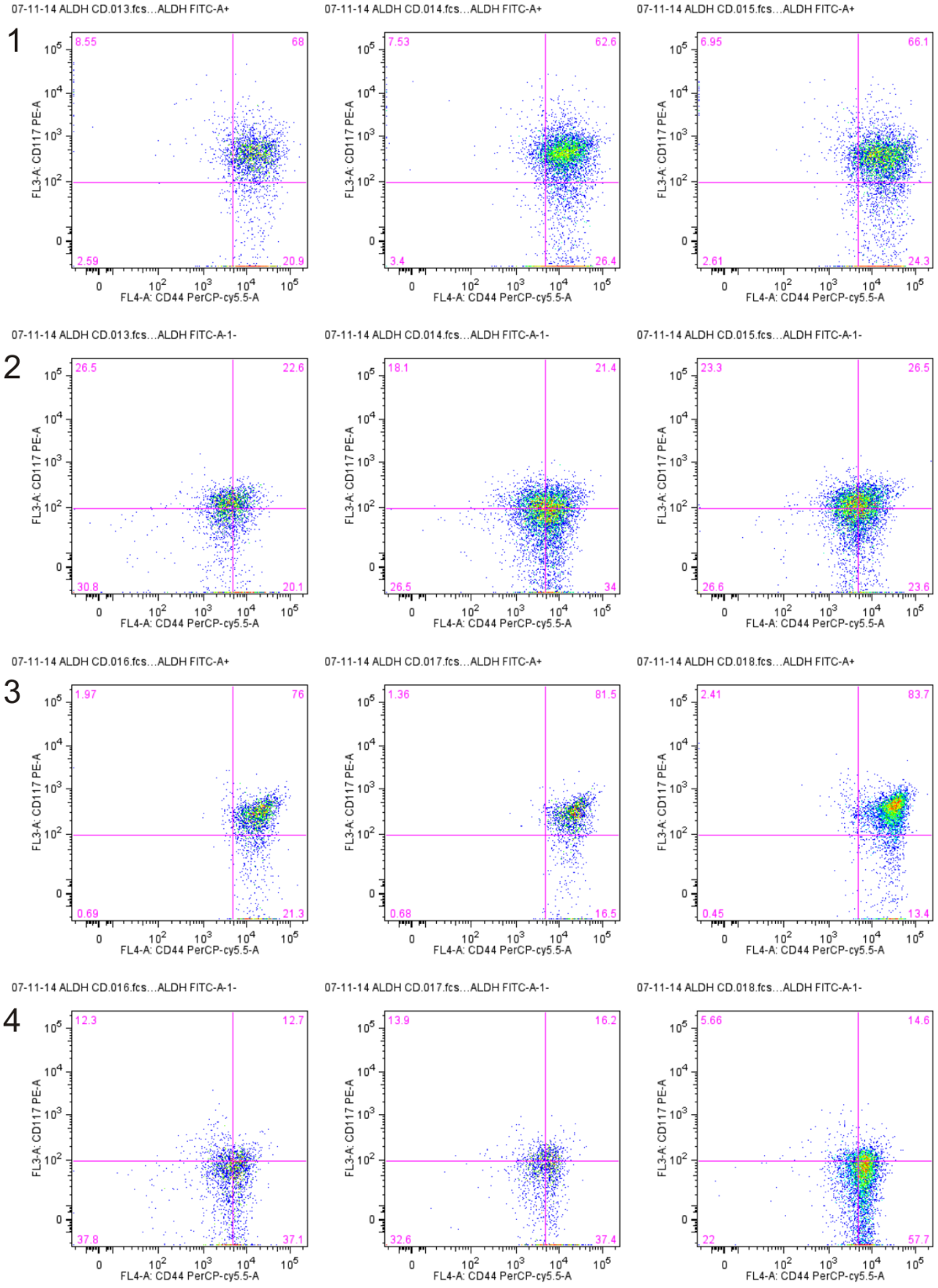
Supplementary figure 3 CD90 expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



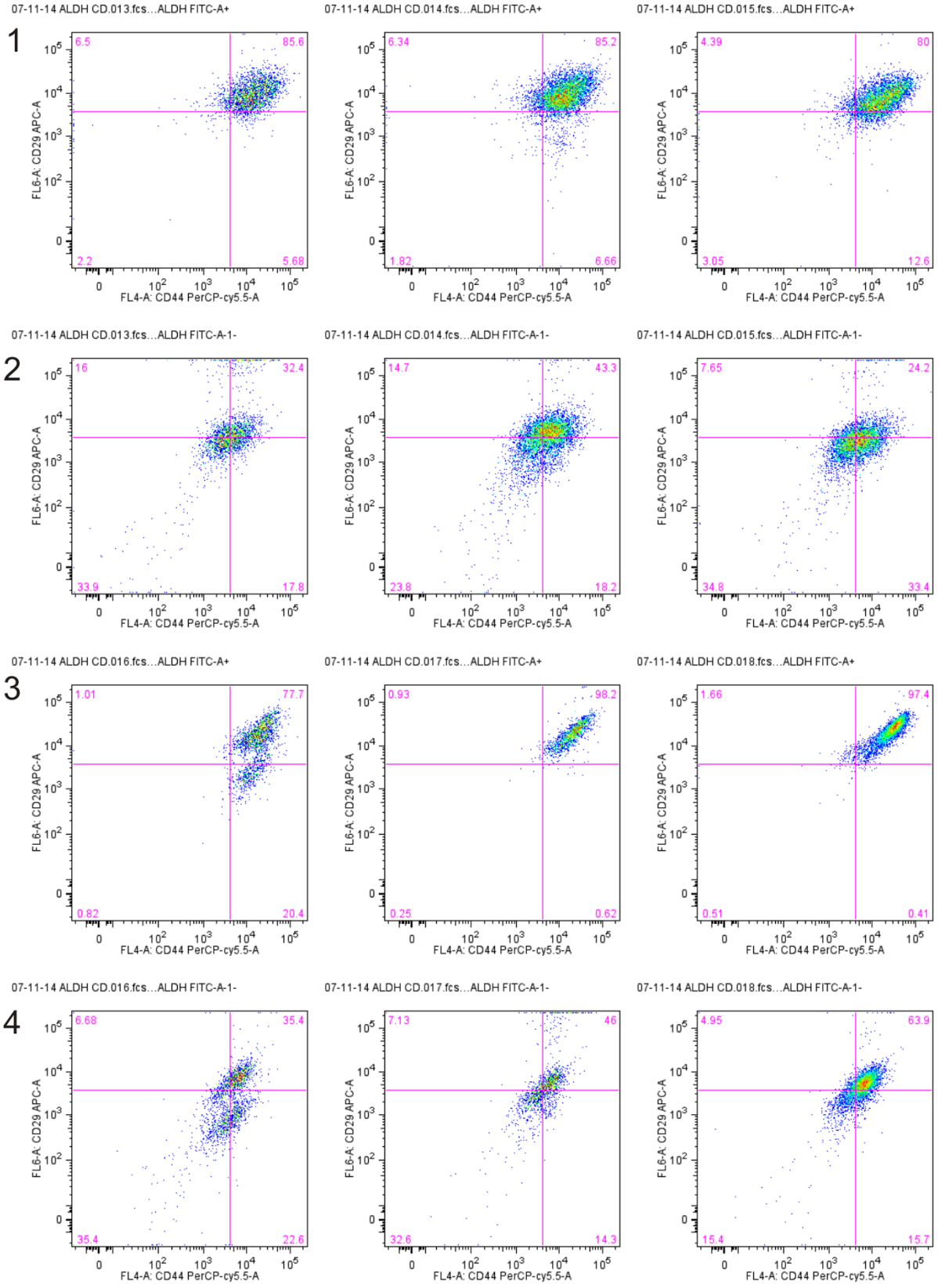
Supplementary figure 4 CD133 expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



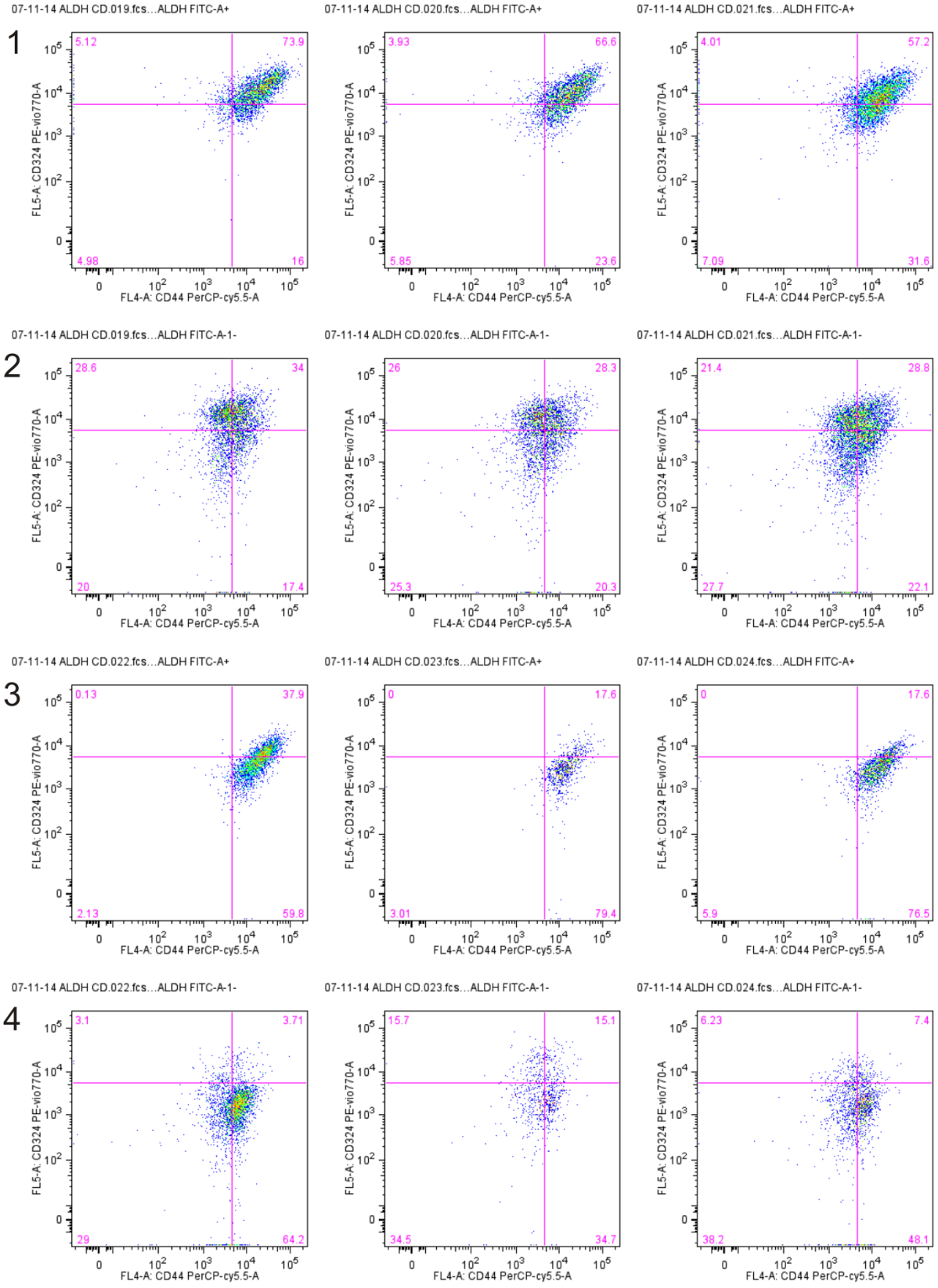
Supplementary figure 5 CD117 expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



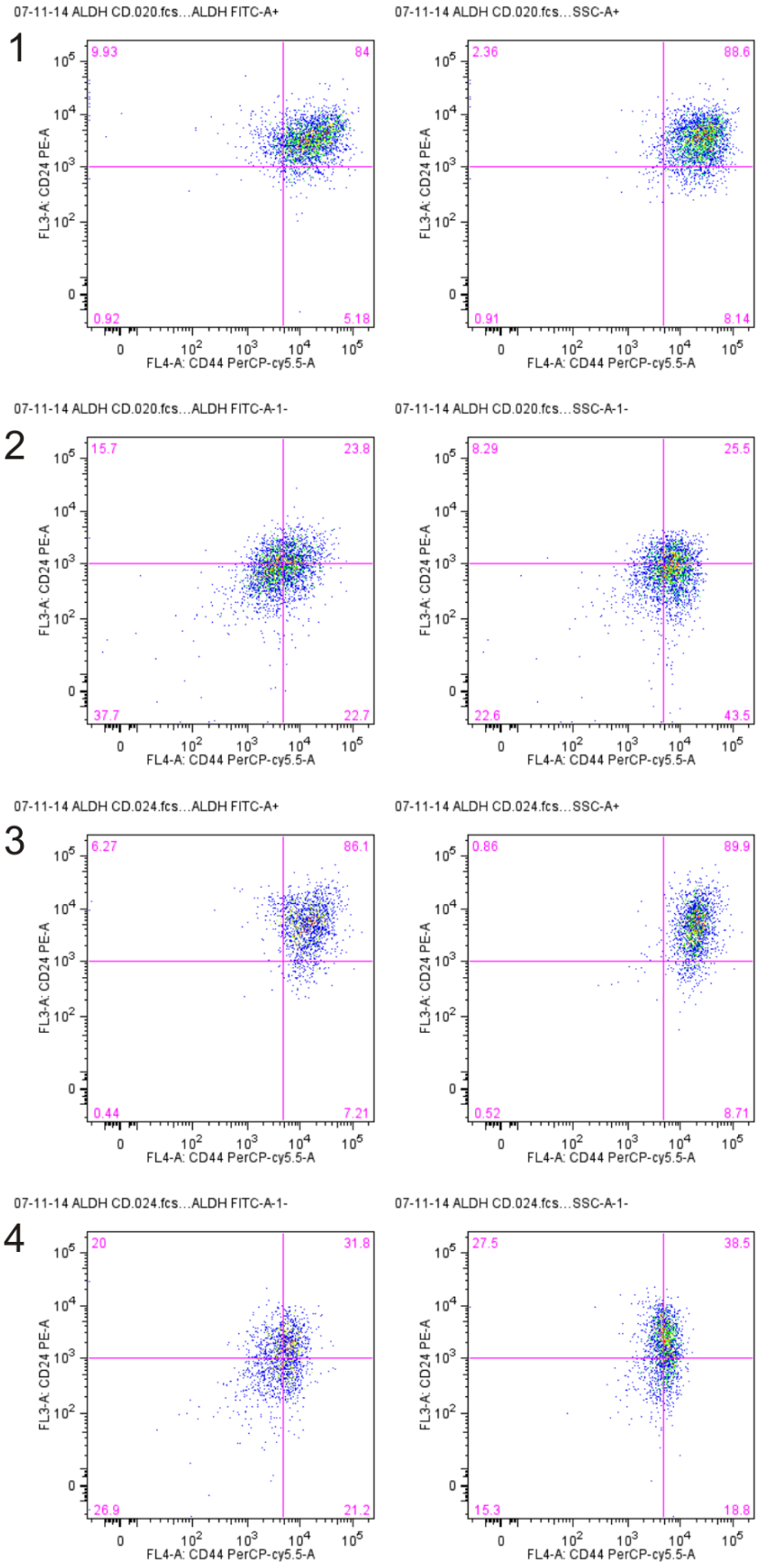
Supplementary figure 6 CD29 expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



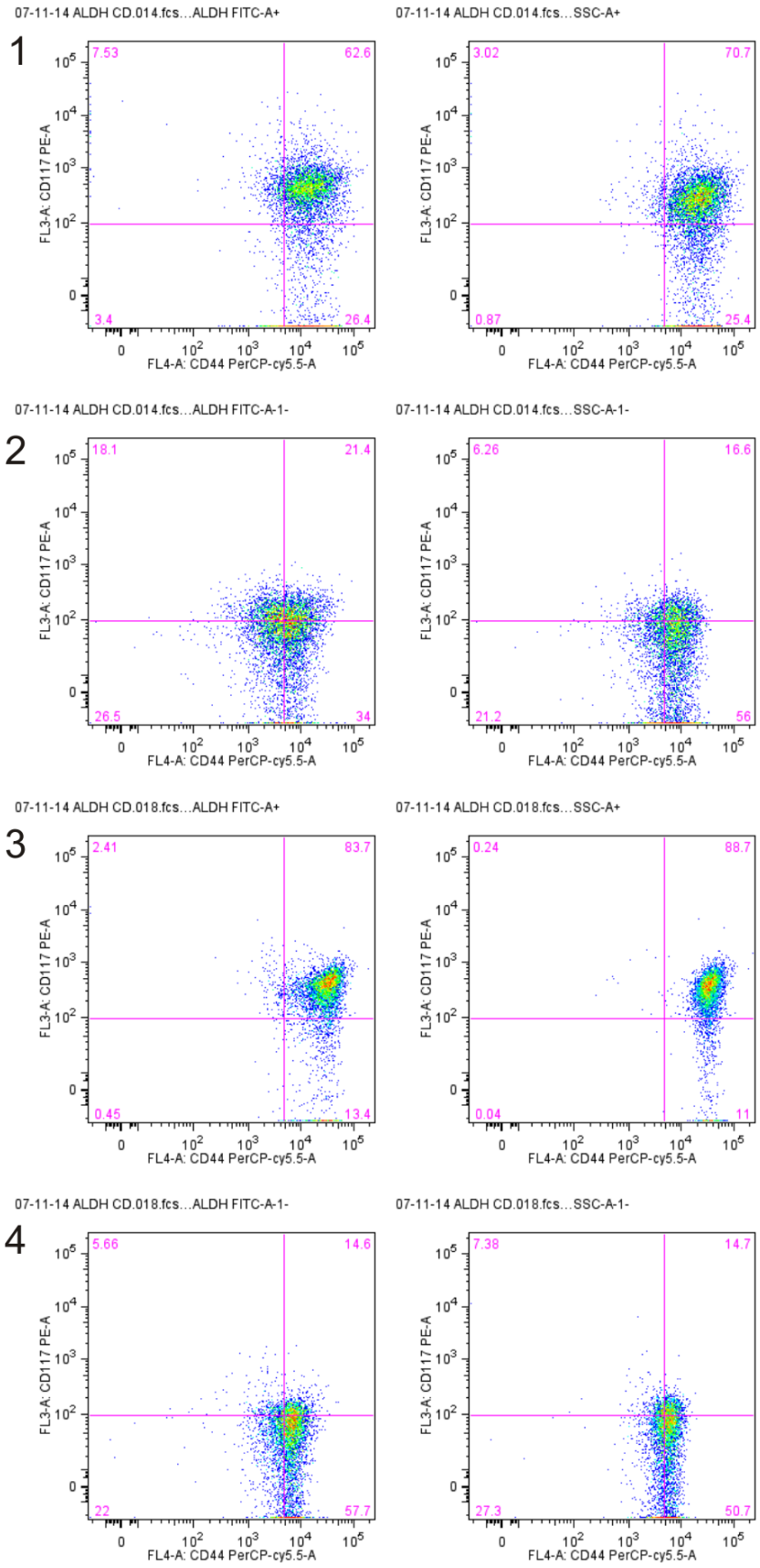
Supplementary figure 7 CD324 (E-Cad) expression in 5% ALDH high and low D492 and D492M cells

1: ALDH high D492. **2:** ALDH low D492. **3:** ALDH high D492M. **4:** ALDH low D492M.



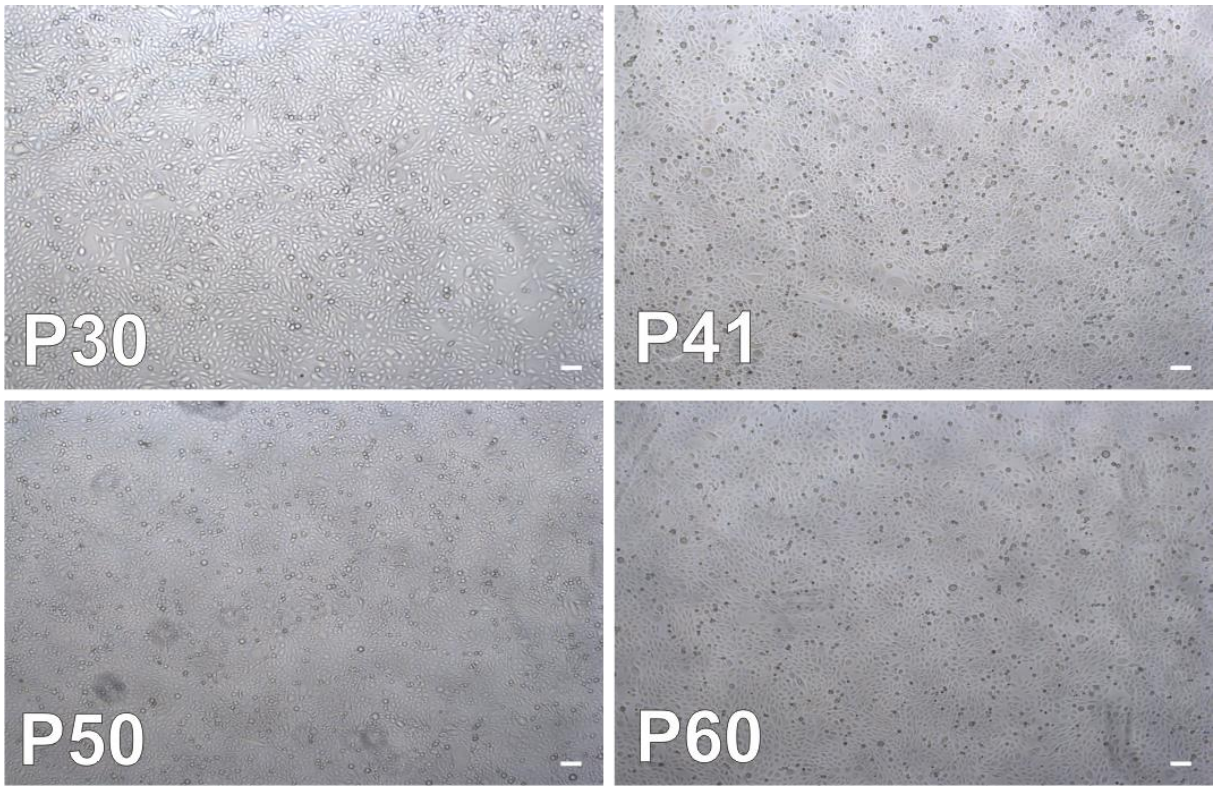
Supplementary figure 8 Comparison of ALDH and SSC in D492 and D492M – CD24

1: ALDH high and SSC high in D492. **2:** ALDH low and SSC low in D492. **3:** ALDH high and SSC high in D492M. **4:** ALDH low and SSC low in D492M. ALDH and SSC levels are set around 5%



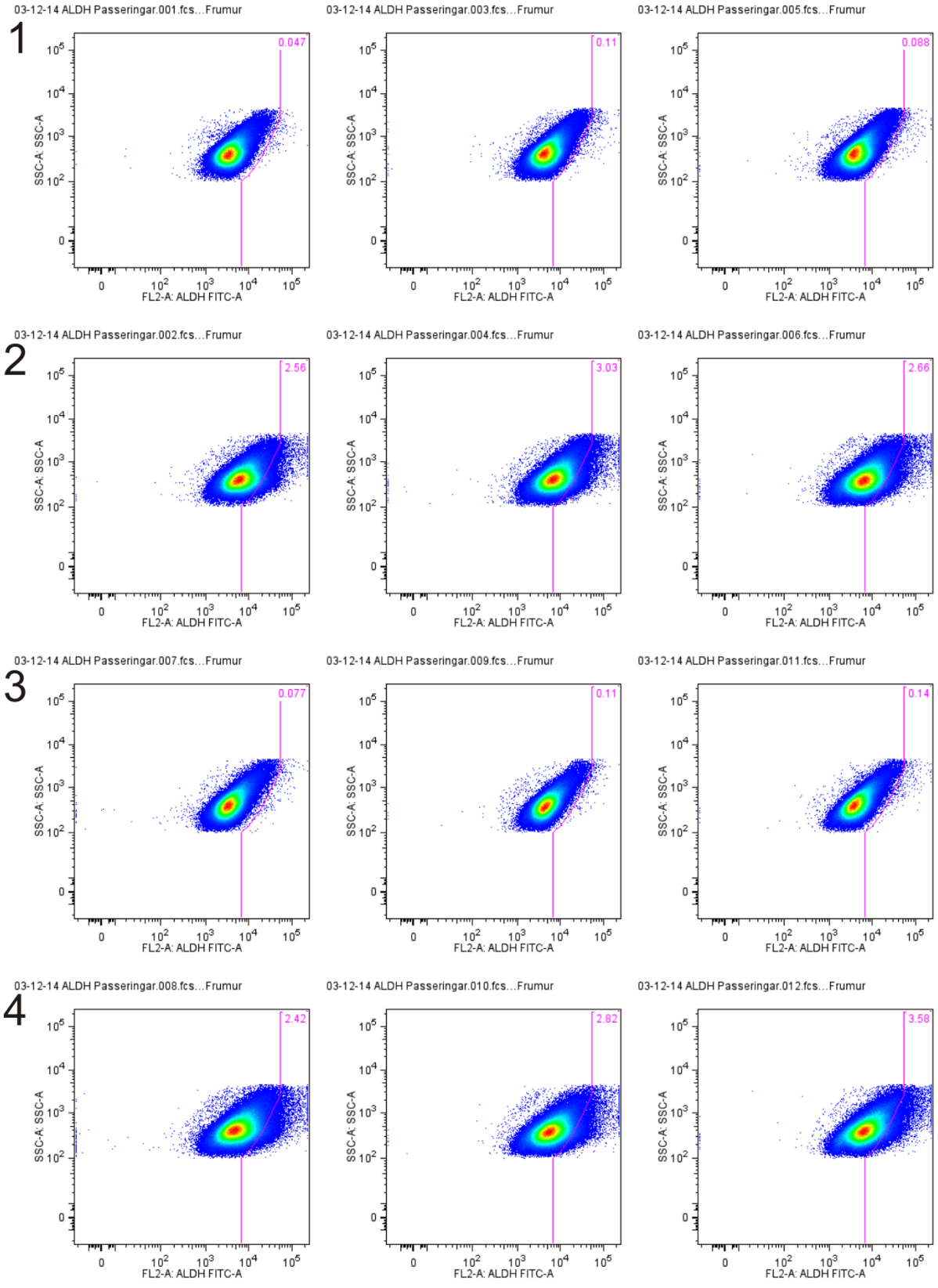
Supplementary figure 9 Comparison of ALDH and SSC in D492 and D492M – CD117

1: ALDH high and SSC high in D492. **2:** ALDH low and SSC low in D492. **3:** ALDH high and SSC high in D492M. **4:** ALDH low and SSC low in D492M. ALDH and SSC levels are set around 5%



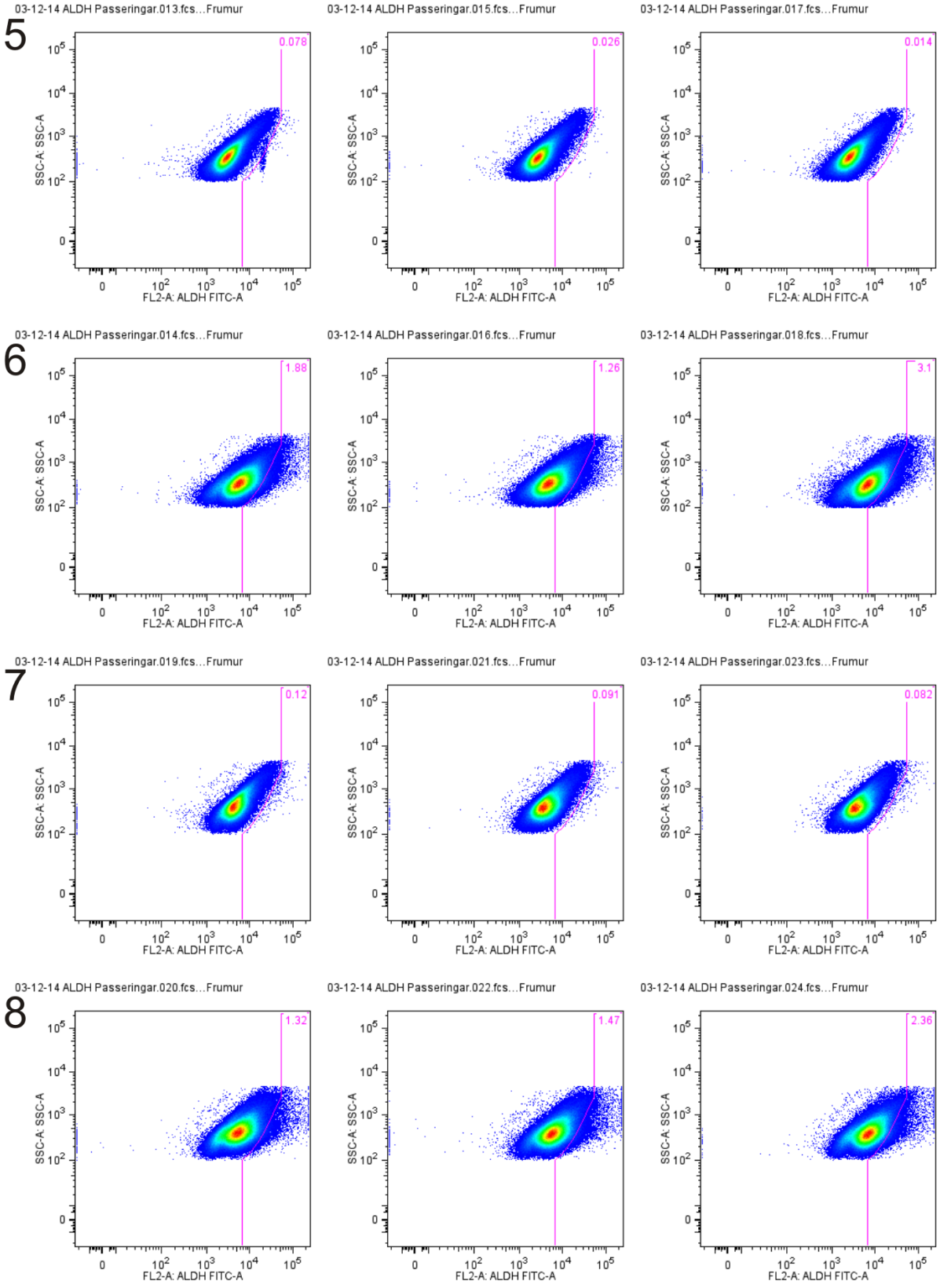
Supplementary figure 10 Different cell passages of D492

492 cells that have been passaged 30, 41, 50 and 60 times. Phase contrast light microscopy, original magnification 50x, bar – 100 µm.



Supplementary figure 11 ALDH activity in cell passages 30 and 41

1: Negative controls for P30. **2:** P30 samples **3:** Negative control for P41. **4:** P41 samples



Supplementary figure 12 ALDH activity in passages 50 and 60

5: Negative controls for P50. **6:** P50 samples **7:** Negative control for P60. **8:** P60 samples

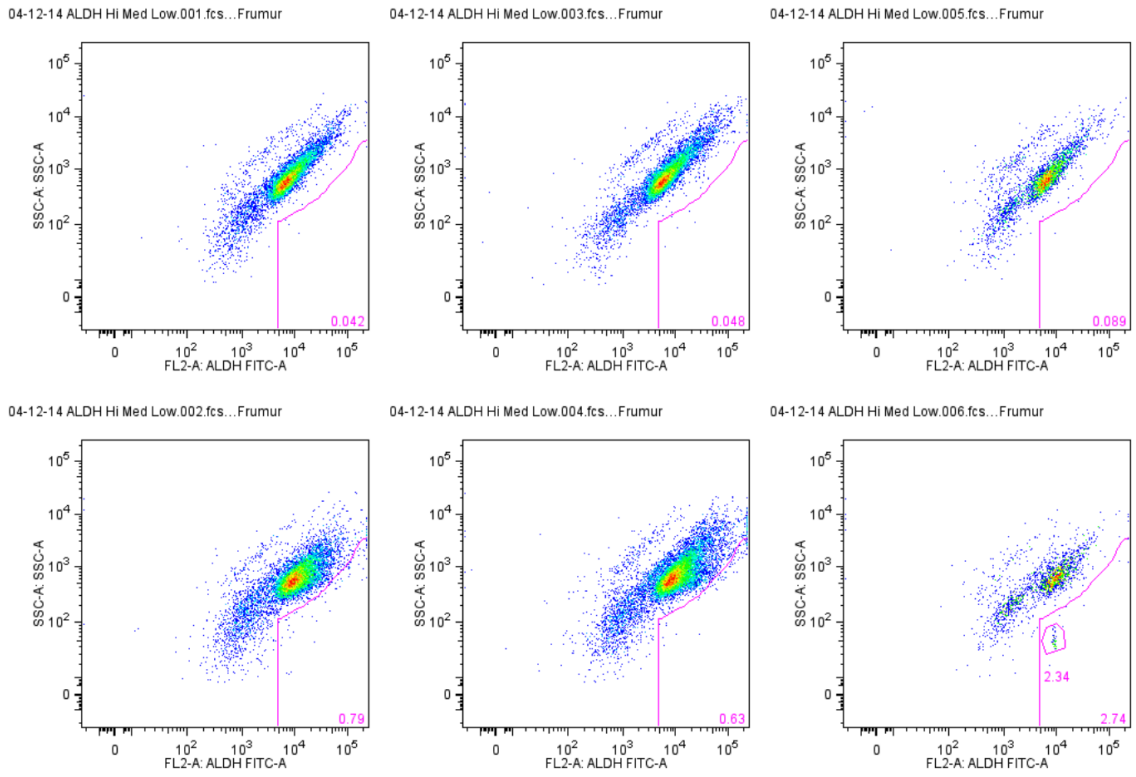
7.3 ALDH activity post sorting – density of 1000 and 2000 cells

ALDH sorted fractions were seeded at different density (1000, 2000 and 5000) and grown for 10 days on EGM5 medium. Seeding of 5000 cells gave the most relevant results with variance in ALDH activity in ALDH sorted fractions. Following are the results for 1000 and 2000 cells seeded.

7.3.1 1000 cells seeded

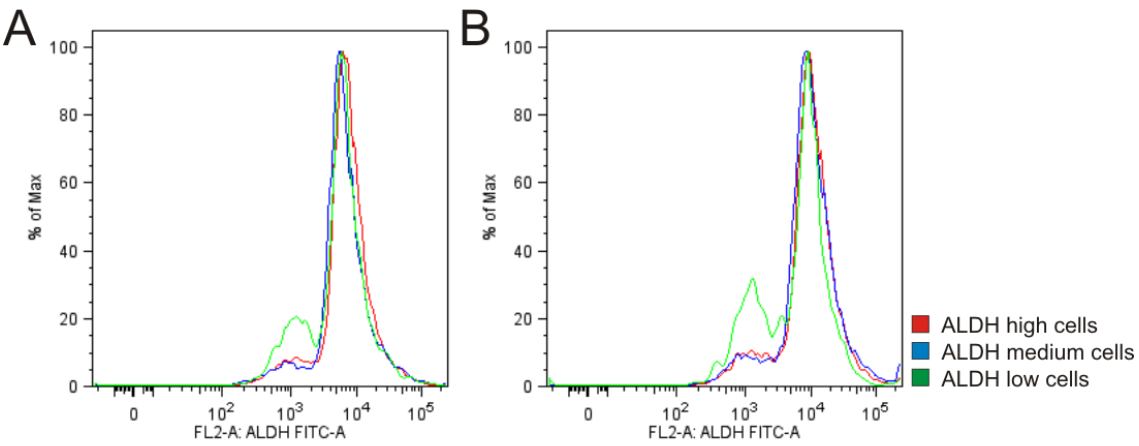
For 1000 cells seeded, ALDH activity was similar for ALDH high and medium cells (Supplementary figure 13). The activity seems higher in ALDH low albeit the cells contributing mostly have abnormal SSC properties and are likely an artifact (gated – 2.34%).

Combined histograms for ALDH stained populations of negative controls (Supplementary figure 14A) and samples (supplementary figure 14B). For control and sample of ALDH low cells a low peak at the left side of the curve stands out compared to the other curves. The ALDH activity appears though similar between sorted cell fractions.



Supplementary figure 13 FACS analysis of ALDH activity in ALDH sorted high, medium and low cells – 1000

ALDH sorted populations, from cultures of 1000 cells seeded. **Top row:** Negative control for ALDH high, medium and low cells. **Bottom row:** Samples for ALDH high, medium and low cells matched to the negative controls above.



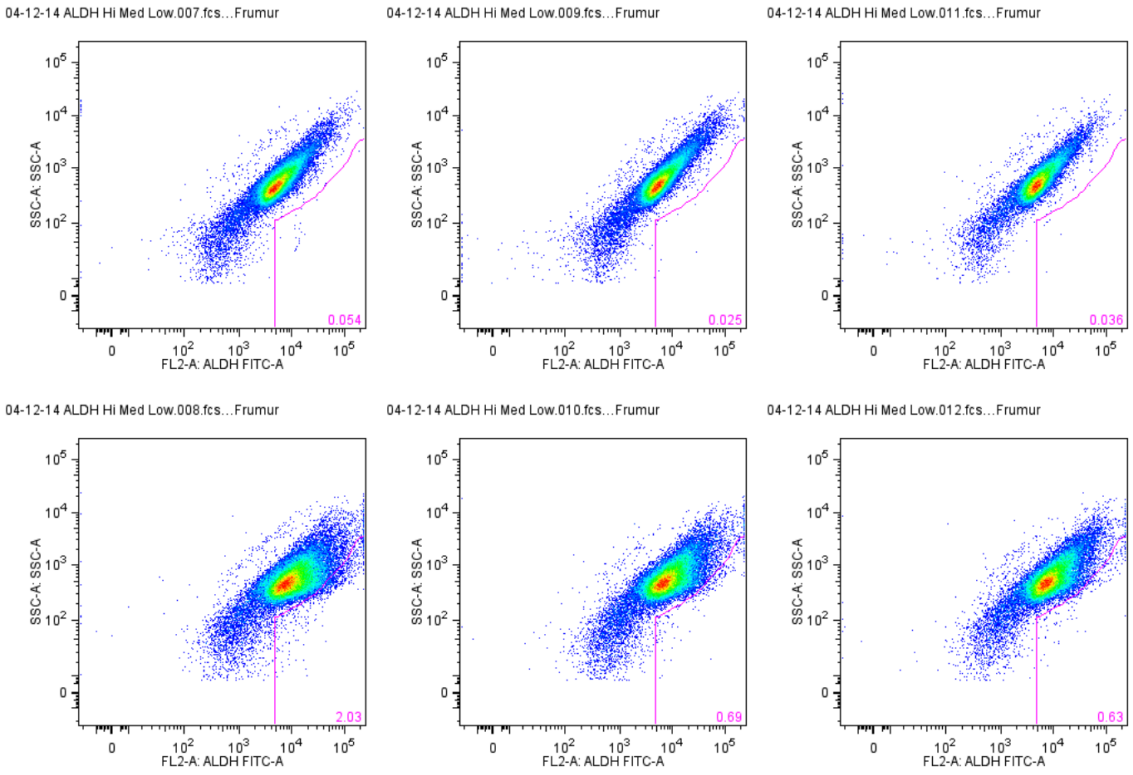
Supplementary figure 14 ALDH activity in 1000 cells seeded – histogram curves combined

A: ALDH histograms for negative controls. **B:** ALDH histograms for samples.

7.3.2 2000 cells seeded

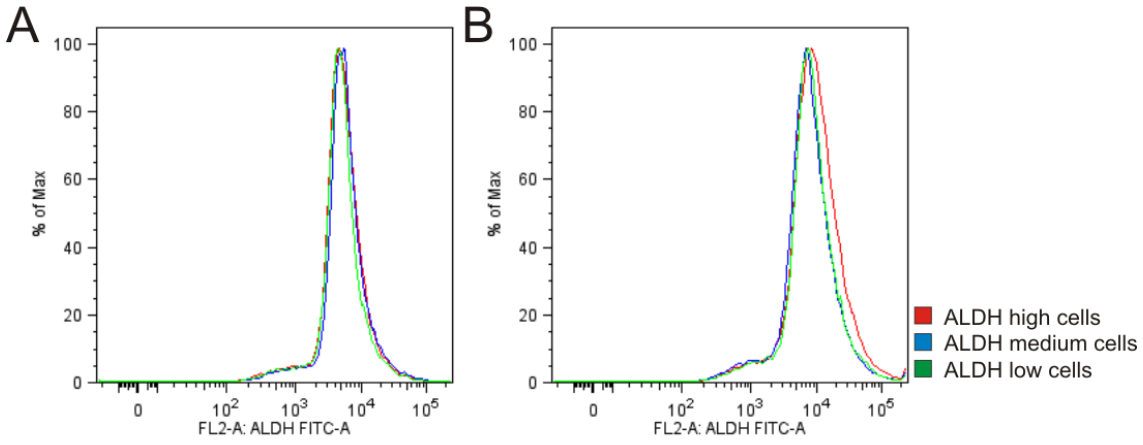
For 2000 cells seeded ALDH activity is similar for ALDH medium and low cells while higher in ALDH high cells (Supplementary figure 15).

ALDH histograms in negative controls are mostly in agreement (Supplementary figure 16A) while samples show an increase in ALDH high fraction (Supplementary figure 16B).



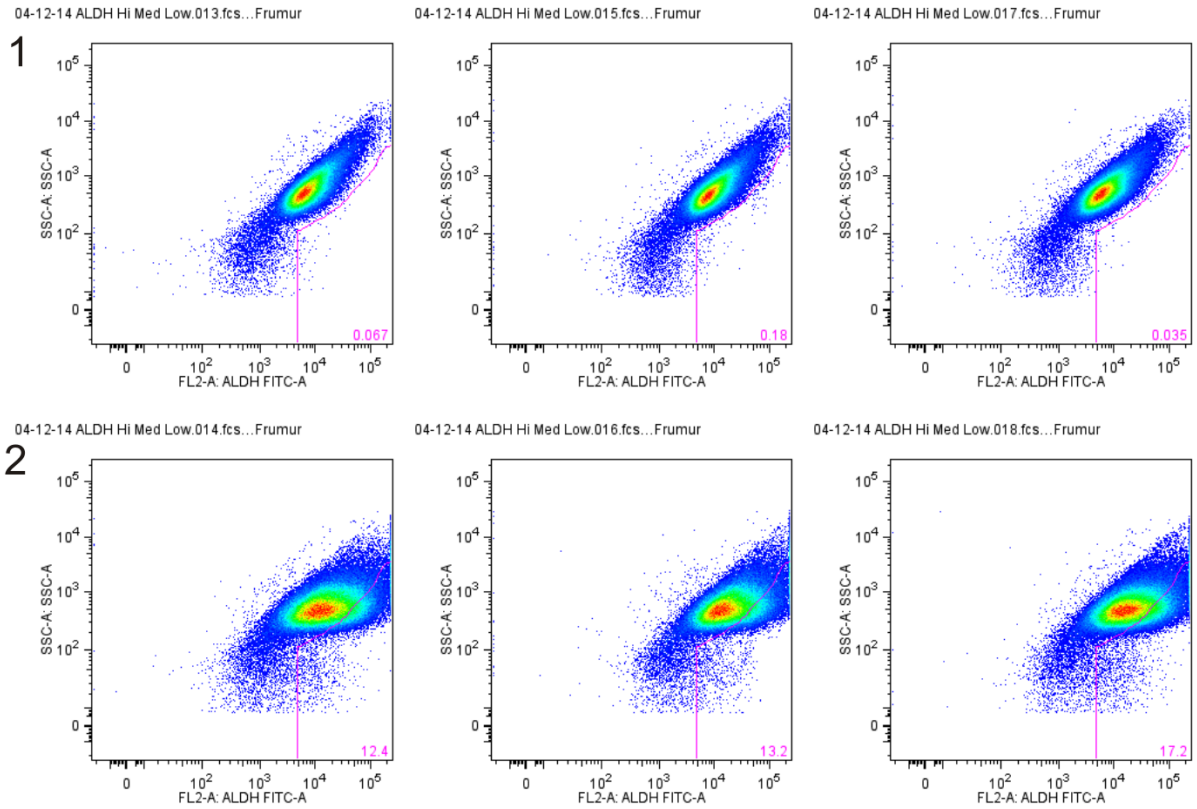
Supplementary figure 15 FACS analysis of ALDH activity in ALDH sorted high, medium and low cells – 2000

ALDH sorted populations, from cultures of 2000 cells seeded. **Top row:** Negative control for ALDH high, medium and low cells. **Bottom row:** Samples for ALDH high, medium and low cells matched to the negative controls above.



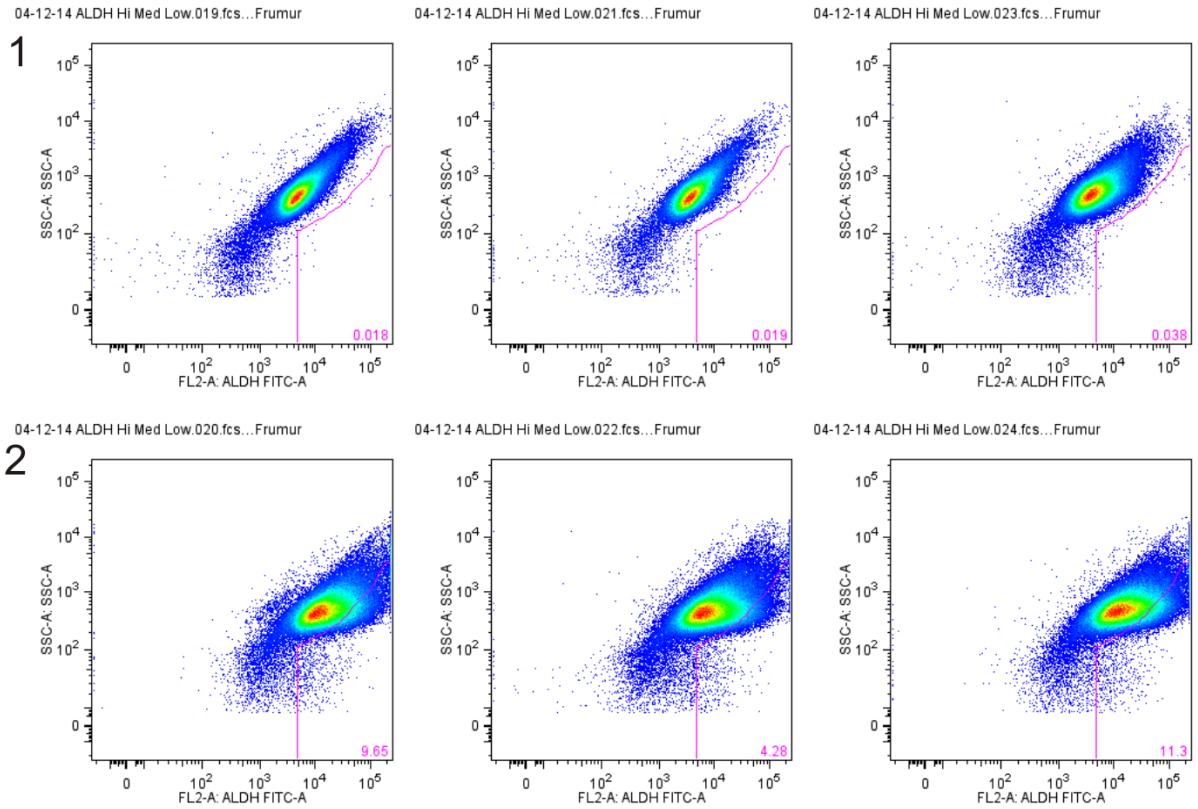
Supplementary figure 16 ALDH activity in 2000 seeded cells – histogram curves combined

A: ALDH Histograms for negative controls. **B:** ALDH histograms for samples.



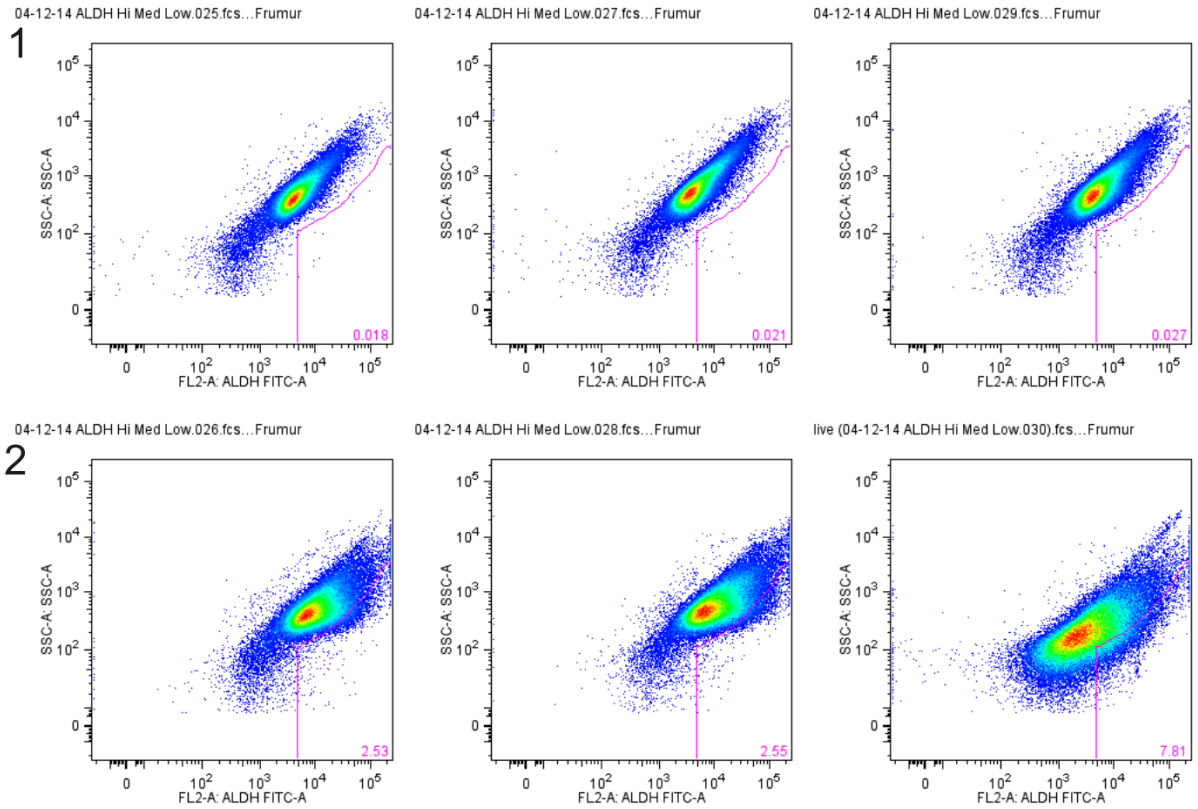
Supplementary figure 17 ALDH activity in ALDH high cells – 5000 cells originally seeded

1: Negative controls for ALDH high cells in triplicate consecutively lined. **2:** Samples for ALDH high cells that respectively correspond to the above negative controls.



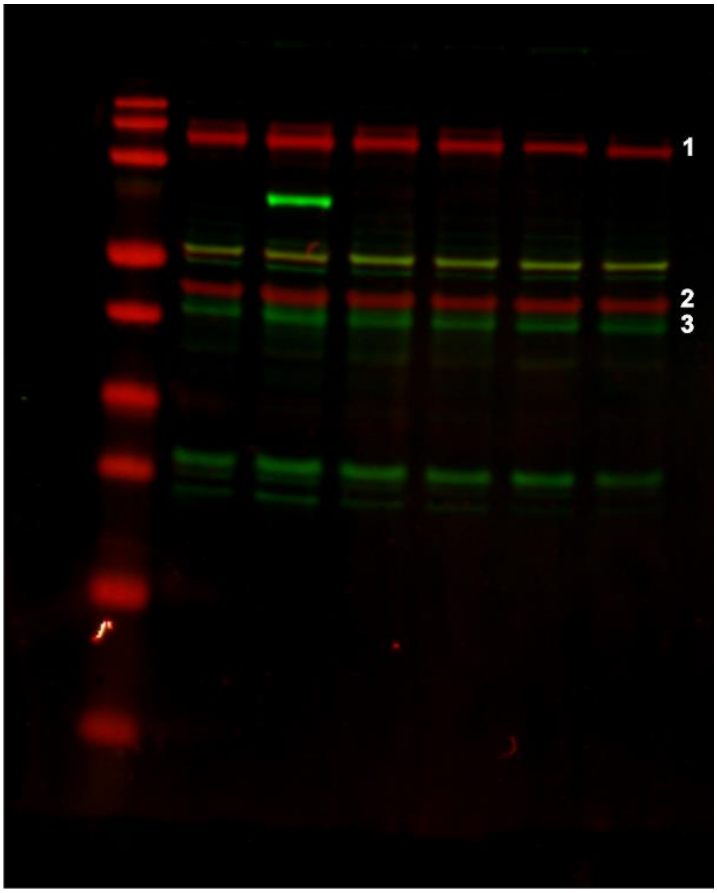
Supplementary figure 18 ALDH activity in ALDH medium cells – 5000 cells originally seeded

1: Negative controls for ALDH medium cells in triplicate consecutively lined. **2:** Samples for ALDH medium cells that respectively correspond to the above negative controls.



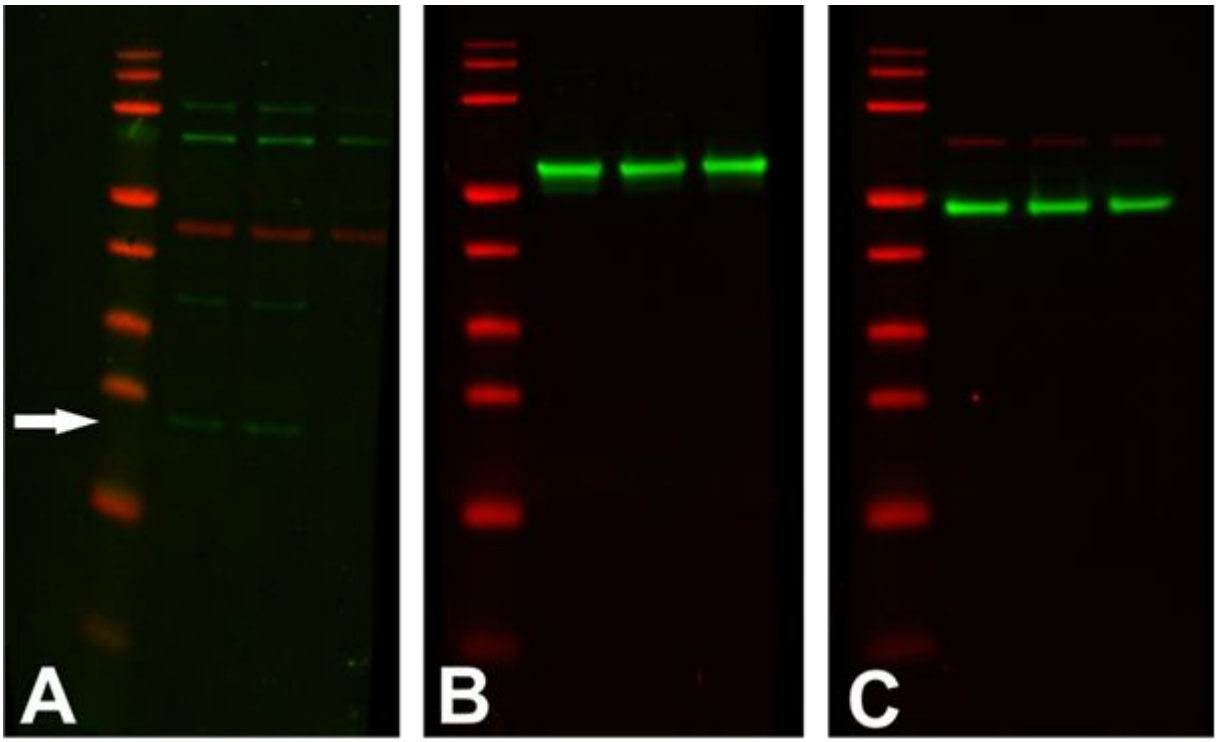
Supplementary figure 19 ALDH activity in ALDH low cells – 5000 cells originally seeded

1: Negative controls for ALDH low cells in triplicate consecutively lined. **2:** Samples for ALDH low cells that respectively correspond to the above negative controls.



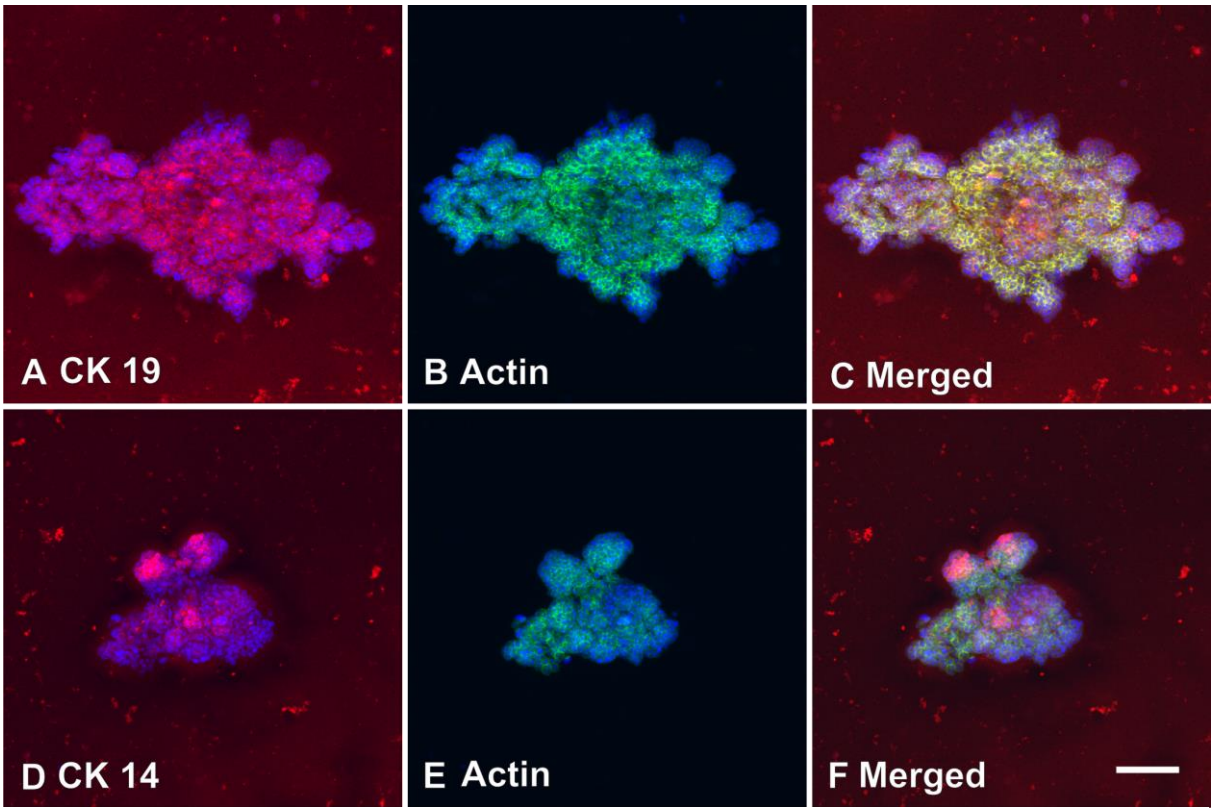
Supplementary figure 20 Original Western blot for E-Cadherin, Actin and EpCAM.

E-Cadherin (1), Actin (2) and EpCAM (3), other green bands belong to EpCAM.



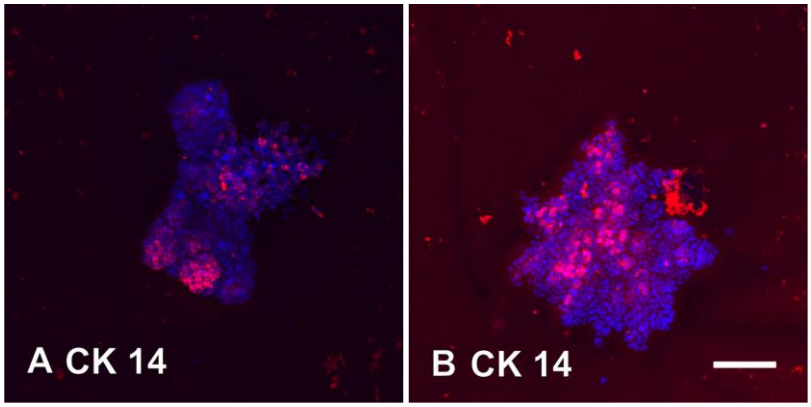
Supplementary figure 21 Original western blot of Actin, Thy-1, CK14, CK5/6 and CK19

A: Red bands stand for Actin and green bands pointed by arrow for Thy-1. Other green bands belong to Thy-1. **B:** Green band stands for CK14. **C:** Red bands stand for CK5/6 and green bands for CK19.



Supplementary figure 22 CK19 and CK14 staining of colonies from ALDH high cells without BRENCs

A-C: A colony from ALDH sorted high cells stained with CK19 and F-actin. **D-F:** A colony from ALDH sorted high cells stained with CK14 and F-actin. CK19/CK14 – red. F-actin – green. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.



Supplementary figure 23 CK14 staining of colonies from ALDH high and low cells with BRENCs.

A: A colony from ALDH sorted high cells stained with CK14. **B:** A colony from ALDH sorted low cells stained with CK14. CK14 – red. Nuclear counterstaining – blue (DAPI). Bar – 100 µm.