Refinement of Flow Cytometry Based Methods for Identification and Characterization of Differentiation Specific Intestinal Epithelial Cells

Author:

Marina Elmelund Sørensen

Supervisors:

Karen Dybkær

Alexander Schmitz

Laurids Østergaard Poulsen

Ursula Falkmer

Contents

Resumé

Tyk- og endetarmskræft er på verdensplan den 3. mest hyppige kræftform, hvor kun omkring halvdelen af de kræftramte overlever sygdommen i mere end 5 år. I forhold til prognosen og med henblik på mere målrettede behandlingstilbud, vurderes det at have stor klinisk relevans at benytte genekspressionsanalyser til at bestemme hvilken del af tarmslimhinden som kræften er opstået i. Målet med det overordnede projekt er at kunne afgøre om en tumor er opstået i stamceller i bunden af en tarmkrypt eller i epitelvæv i toppen af en tarmkrypt. Tidligere forskning har haft fokus på karakterisering af cellerne i kræftsvulster hvorimod der i dette studie tages udgangspunkt i normalt væv for at undersøge generelle metoder til klassifikation baseret på gensignaturer. I dette studie blev flowcytometri evalueret som metode til at identificere de relevante celler i en suspension af celler fra normalt tyktarmsvæv, der blev indsamlet i forbindelse med tarmresektion på Aalborg Universitetshospital. Forsøget blev godkendt af den lokale etiske komité og alle patienter gav informeret samtykke til medvirken i forsøget.

Flowcytometriske analyser blev udført på celler i suspension med en række forskellige differentieringsspecifikke overfladeantistoffer med henblik på at kunne identificere celler med bestemte egenskaber. Selektionskriterierne blev opsat til sortering (gating) ud fra parametre, der afspejler cellestørrelse og granularitet. Det viste sig svært at opnå entydige resultater på grund af ikke-specifik cellebinding til de testede antistoffer. Det blev valideret gennem en række forsøg, at cellesuspensionerne havde et stort indhold af døde eller døende celler. Ud fra en række eksperimenter anbefales det at bruge antistoffet FVD, der binder til celler uden intakt cellemembran, til frasortering af disse celler for at reducere den ikke-specifikke binding. Gennem forsøg med titrering af udvalgte antistoffer viste det sig fordelagtigt at optimere på antistofkoncentrationen. De flowcytometriske undersøgelser demonstrerede, at Lgr5 der er et af de centrale antistoffer i forbindelse med identifikation af de relevante stamceller, også bandt sig på leukocytter. Dette var ikke forventeligt og det anbefales derfor at gentage forsøget under mere kontrollerede forhold. Magnetisk-aktiveret cellesortering blev med stor succes afprøvet som metode til at øge koncentrationen af EpCAM positive celler i en suspension af en HT-29 cellelinje. Det anbefales, at dette forsøg udføres på en suspension af celler fra normalt tarmvæv sammen med antistoffet Lgr5. De grundlæggende metoder til celleidentifikation blev undersøgt ved hjælp af flowcytometri. En række forslag til optimering af metoden blev identificeret, og det anbefales derfor at udføre en række nye eksperimenter for at afklare om identifikation ved hjælp af differentieringsspecifikke antistoffer er mulig.

Abstract

Colorectal cancer is the third-most prevalent cancer type with a 5-year survival rate around 50%. It has been hypothesized that it has clinical relevance to identify whether the colon cancer cells originate from the apical epithelial cells or the basal stem cells of the intestinal crypts. The aim of the present study was to assess the initial methods for identifying the cells of interest using multiparametric flow cytometry on single cell suspensions from normal colonic epithelial tissue stained with a panel of differentiation-specific antibodies. The results of the single-stain experiment were inconclusive due to unspecific staining caused partly by the presence of a high number of non-viable cells in the cell suspensions. This was documented in a viability stain experiment on a human colorectal adenocarcinoma cell line (HT-29) where viability staining using FVD was found to be recommended in future studies. In a series of antibody titration experiments using selected antibodies it was demonstrated that it is important to establish the optimal antibody concentration to reduce the risk of unspecific staining. It is also important to further test the validity of the Lgr5 stain since it was found to stain cell populations that were not expected to express Lgr5. The principle of magnetic-activated cell sorting (MACS) was successfully tested on a HT-29 cell line where it was possible to achieve a significant enrichment of EpCAM positive cells. This principle should be further investigated in a suspension from normal colonic epithelial tissue using Lgr5 antibody. Based on the important key findings of the performed experiments it is recommended to do additional experimentation to be able to draw a conclusion on the main hypothesis.

Introduction

Colorectal cancer (CRC) is a potentially life-threatening disease with a world-wide incidence of 1.360 million cases, corresponding to 9.7% of all cancer cases (1). The number of CRC-related deaths is 694,000 world-wide, corresponding to 8.5% of all cancer deaths (1). The prevalence of CRC in Denmark is around 30,000 with an incidence of 5,000 annual cases (2). About half of the new cases occur in the age group from 55-75 years while 40% of cases occur in people over 75 years of age (3). After 5 years, the number of CRC survivors is around 50% (2). The treatment of CRC is typically based on surgery in combination with chemo- and radiotherapy (2).

According to the Danish Colorectal Cancer Group (DCCG), classification of CRC tumors should be done according to the anatomic Tumor, Node, and Metastasis (TNM) method, where the tumor is assessed according to tumor size and penetration of the colonic wall (T), involvement of lymph nodes (N), and metastatic tumor spread (M) (4).

Histological classification of the tumors is done into 9 different categories (5) with glandular adenocarcinomas being the most common type encountered in 77% of cases (6). Further analysis is typically performed to assess whether the glandular adenocarcinoma has a low degree of differentiation (7). If so, the prognosis is usually negatively affected (8).

The molecular mechanisms of colorectal oncogenesis have been studied extensively (9). One of the strategies for developing molecular classifications of CRC is based on "gene expression profiling" (GEP) of tumor biopsies prior to treatment. Based on the mRNA expression levels, the CRC tumors are stratified into different risk groups where the impact on outcome is assessed in statistical analysis. However, this has so far not been convincing. reproducibility between studies in the identification and characterization of GEP defined molecular subgroups. In an attempt to stratify colon cancer, various gene expression based classification systems have been reported. Six of those were coalesced by an international consortium into four consensus subtypes: Microsatellite instability immune (CMS1), canonical (CMS2), metabolic (CMS3) and mesenchymal (CMS4) (9). In CMS2, there is a "cluster" of genes in the malignant biopsies that are characteristic of basal crypt-like tissue, whereas CMS3 contains an apical crypt –like "cluster" of genes (10). The CMS2 type is characterized by high proliferation rate and is associated with decreased survival while the CMS3 type is characterized by low proliferation and low stromal invasion (10)

CRC is a malignancy of the epithelial type usually derived from normal colonic mucosa through the adenoma stadium (11). The normal colonic epithelial lining houses a subset of intestinal stem cells (ISCs), as shown in figure 1.1, responsible for normal homeostatic production of differentiated mature cells (12). It has a very high turnover rate which permits the replacement of the entire intestinal epithelial lining every 5 to 7 days (13). ISCs are located at the base of the gastrointestinal crypts (13)(14). Maturing ISCs develop to transit-amplifying daughter cells that after 4 to 5 divisions migrate upwards along the crypt and become fully matured epithelial cells of different type at the apical part of the crypt (13).

The focus of this study has been on identifying and isolating the fast-cycling ISCs that are located at the base of the gastrointestinal crypts (13)(14). Despite a well-established understanding of ISCs role in homeostasis and injury, a comprehensive and well-founded overview of their molecular markers, that allows functional use within diagnostics, treatment, and prognostics, is still missing (13). Therefore, further investigation of ISC pathways and their molecular expression is very important.

Markers for normal differentiation include surface expressed proteins with LGR5 being specifically expressed on the stem cells located in the basal crypt (15). LGR5 is a target gene of the Wnt-signaling pathway which is believed to be a main player in ISCs proliferation (13). See the graphical presentation of the LGR5 expression through the intestinal crypt in fig. 1.1. Other proteins co-expressed on Lgr5 positive cells are CD44 and CD166 (13). See the overview of the previously described differentiation specific markers in table 1.1

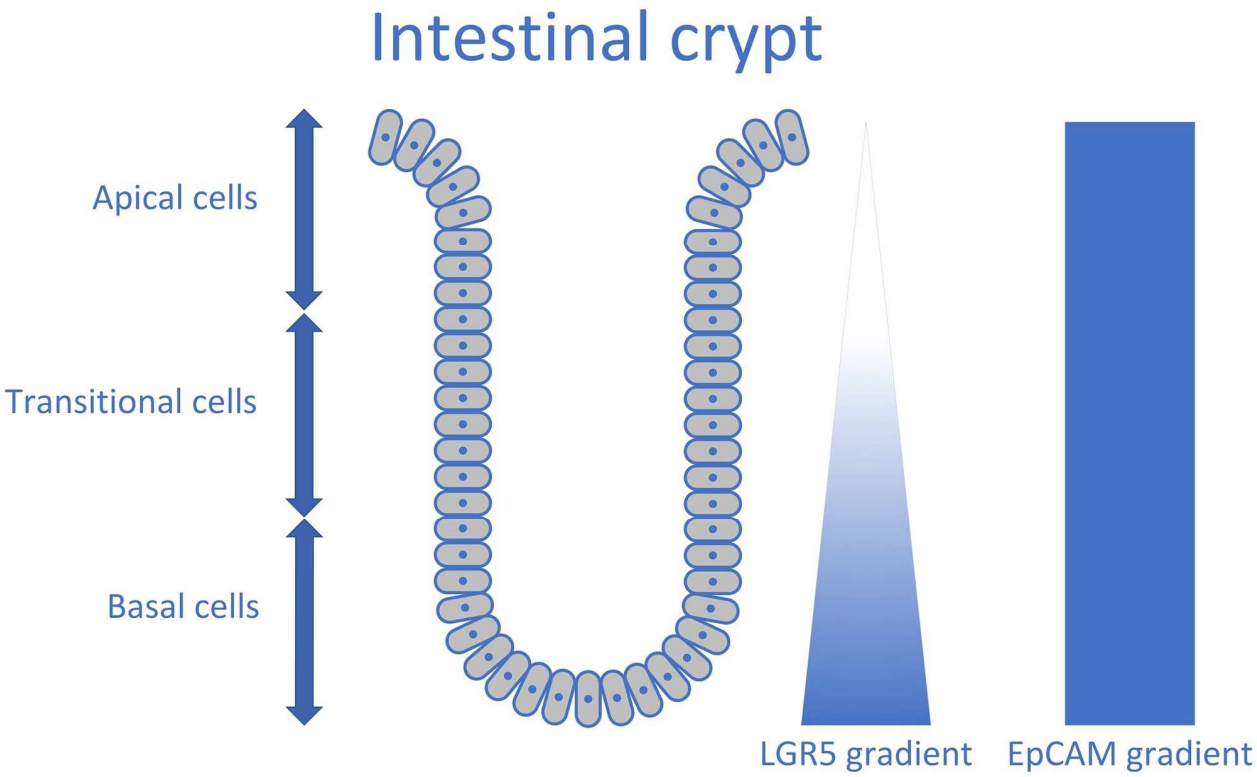


Figure 1.1 Schematic representation of the intestinal crypt. Intestinal epithelial cells originate from stem cells in the crypt that are expected to express LGR5 in a gradient from the basal to the apical part of the intestinal crypt. EpCAM is a universal biomarker of epithelial cells that is expected to show a uniform distribution throughout the crypt.

Colorectal cancer consists of abnormal and heterogeneous tissue that originates from a subset of cells referred to as cancer stem cells (CSC). CSCs are described as having stem cell-like properties, such as capacity for self-renewal and differentiation to multiple cell lineages. (16)(14). CSCs are viewed as long-lived, low proliferative, cytotoxicity-resistant cells capable of giving rise to subsets of actively proliferating progenies just like normal stem cells give rise to functional physiological units (17). Together, these cell type subsets are capable of tumor formation, propagation and maintenance (17)

The concept of CSCs was initially introduced and proven in the field of acute myeloid leukemia, where the CSCs were isolated and transplanted between immuno-suppressed mice documenting phenotypic and genotypic regrowth of the original malignant tumor free populations. Later, similar concepts have been documented in various solid tumors including colon tumors (16)(17). It is now believed that eradication of CSCs may lead to the elimination of the entire tumor (16). Various therapies targeting CSCs have been proposed and a panel of specific antibodies applied to identify them. Many surface proteins have been proposed and studied, and it is striking that many of those can be found on the surface of normal adult stem cells (16). This similarity may suggest that CSC might be originating from the normal ISCs through the accumulation of epigenetic and genetic alterations (16). An overview of biomarkers being investigated in relation to CSC research in normal and cancerous tissue is shown in table 1.1. A full overview with references is found in appendix A.

Table 1.1: An overview of biomarkers being investigated in relation to CSC research

Biomarker	Definition	Presumed function in normal tissue	Presumed function in CRC cancerogenesis	Further details
Lgr5	Leucine-rich repeat-containing receptor; member of the G-protein coupled, 7-transmembrane receptor superfamily	An important role in the formation and maintenance of adult ISCs during postembryonic development through Wnt signaling pathway	Genetic transformation and tumor initiation based on mutation of the Wnt signaling pathway and APC/ p53-deficiency.	Lgr5 positive cells are found in a lesser amount of 14-16 per crypt in the gastrointestinal epithelial lining with a gradient through the crypt to the villus axis with the highest expression found at the base.
EpCAM	The epithelial cell adhesion molecule also known as CD326.; cell surface glycoprotein found in simple, pseudo-stratified and transitional epithelia in most organs, glands and neoplasms derived from those	Morphoregulatory role; is involved in proliferation, migration, adhesion, differentiation, and cell signaling in normal epithelia, stem and progenitor cells	EpCAM expression is enhanced in both colon adenomas and CRC; The exact role in the cancerogenesis is unknown. Interaction with Wnt signaling pathways has been proposed.	EpCAM can be observed by IHC in all intestinal epithelium tissue.
CD44	A surface protein also referred as P-glycoprotein, largely distributed throughout the body	Maintenance of hyaluronic metabolism, activation of lymphocytes, and release of cytokines.	Several variants of the protein are found in malignant tissues; Multiple signaling pathways may be involved leading to cellular proliferation, invasion and chemoresistance.	Important isoforms: CD44v3 (associated with poor prognosis in CRC being able to initiate invasion and resist apoptosis), CD44v6 (association to higher metastatic activity and lower disease-free survival within CRC). Clone IM7 has been reported to recognize an epitope common to alloantigens and all isoforms of CD44 that is located between amino acids 145 and 186.
CD324	A marker of e-cadherin, an important cell to cell adhesion molecule in epithelial tissues; Also known as CDH1	Is involved in arranging the epithelial cells in sheets	The loss of cadherins might result in loss of cancer cell adhesiveness, invasion through tissues and metastatic processes	A correlation between low expression of e-cadherin and poor tumor differentiation is detected
CD24	Heat stable antigen B cell-related short glycoprotein located on external membrane of the cell	Promotes cell cycle via increased proliferation and inhibited apoptosis; is a marker of ISCs	The exact role is unknown; Malignant cells expressing CD24 are caracterized by high chemotherapy resistance and self-renewal ability	
CD166	Activated leukocyte cell adhesion molecule; a glycoprotein	Important role in different tissue development during embryogenesis	The exact role in cancerogenesis is unknown	The expression is significantly increased in cancerous processes of various types including CRC.
CD66a	An adhesion molecule of carcinoembryonic family CEACAM and immunoglobulin superfamily; Also known as CEACAM1	Tumor suppressor	The exact role in cancerogenesis is unknown	
EphB2	A member of the receptor tyrosine kinase family	Is responsible for several signaling pathways related to cell growth, migration and sustaining crypt-villus axis in the intestinal crypt	Normal EphB2 levels has been found in the epithelial cells of normal crypts while reduced levels of EphB2 have been identified in CRC which implies importance of its loss in CRC progression	

Based on previous research it is hypothesized that it is possible to differentiate between colon cancer originating from basal stem cells and apical epithelial cells of the colonic crypt based on a panel of differentiation-specific surface markers (18). Based on the hypothesis that gene expression signatures from normal differentiating epithelial cells can be applied onto malignant colon cancer tumors and thereby add a differentiation marker to clinical samples, it becomes possible to test the diagnostic and prognostic impact of such gene signatures. To test this hypothesis, a step-wise approach is planned as listed below.

1) Focus on initial identification and characterization of normal epithelial cells by use of multiparametric flow cytometry.
2) Isolation of ISC located in the basal crypt and mature end-stage epithelial cells located in the apical epithelial crypts by fluorescence activated cell sorting (FACS).
3) Perform gene expression profiling on isolated basal crypt cells and apical epithelial cells to generate differentiation specific gene signatures that can be applied into clinical colon cancer datasets with GEP and outcome data available in online database repositories.

The focus of this study is exclusively on the first aim listed above. The present study is based on previous research performed by department colleagues who wrote the clinical protocol, obtained acceptance from the local ethics committee, tested an optimized protocol for generation of single cell suspensions of normal epithelial cells, and carried out an initial screening of antibodies towards colon cell surface expressed antigens. EpCAM was identified as a reliable marker of the colon epithelial cell population while CD44, CD66a and Lgr5 showed a less distinctive binding pattern through the experiments. Furthermore, EphB2 showed no binding to the colonic epithelial cells.

This study continues the work of identification and characterization of intestinal stem cells collected from healthy colonic tissue with the aim to use the data from healthy tissue as a baseline for comparing to data from malignant tissue. Hereby it is possible to investigate if specific gene signatures have important prognostic value in the treatment of colon cancer.

Methods

A number of experiments were conducted as shown in figure 2.1 to optimize the procedure for identification and characterization of intestinal stem cells collected from healthy colonic tissue.

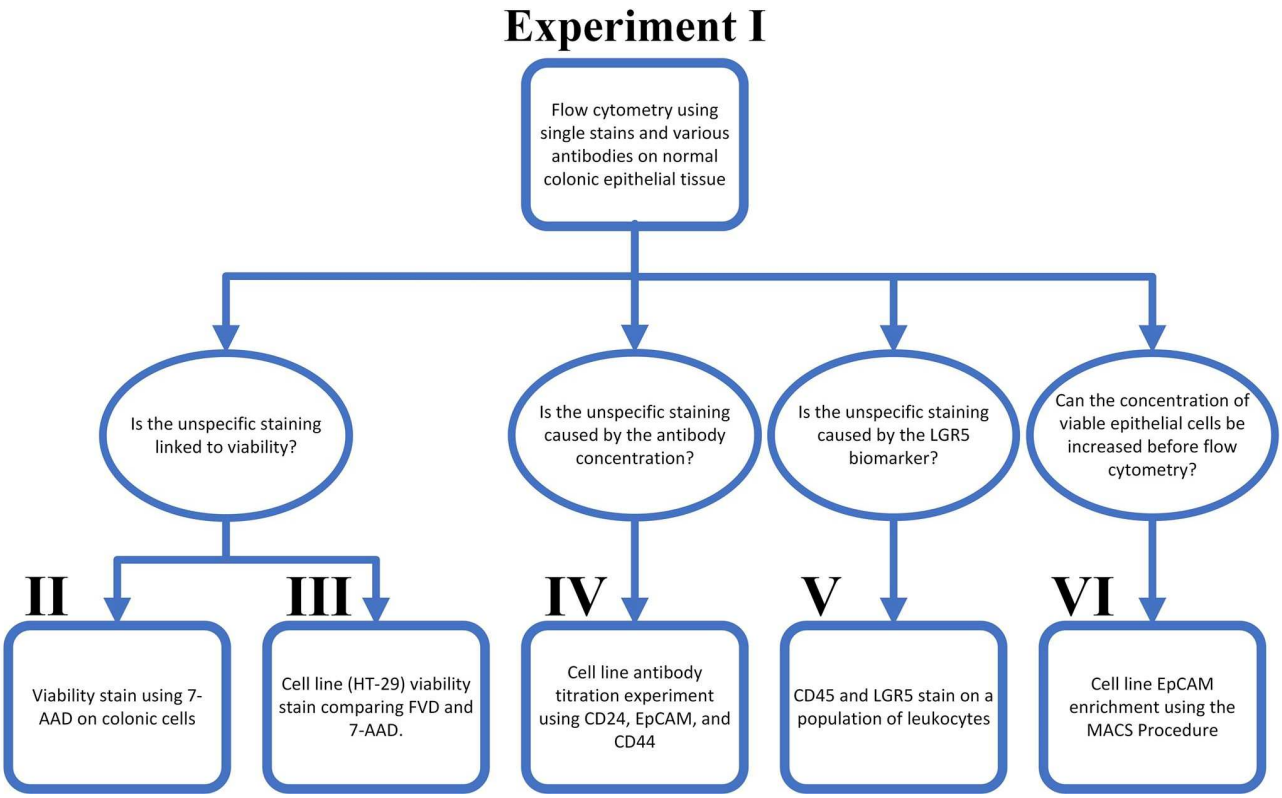


Figure 2.1: The figure provides and overview of the performed experiments. Each rectangle shows an experiment while the ellipses show the question addressed by the experiment. The 6 conducted experiments are numbered in Roman numerals from I to VI for reference.

Selection of differentiation-specific biomarkers

Based on a literature review and previous experimental results from department colleagues, the following markers have been chosen for flow cytometric analysis: LGR5, EpCAM, CD44, CD45, CD324, CD24, CD31, CD166, EphB2, and CD661 (18).

Tissue procurement

Normal colon tissue samples were acquired from resections performed at Aalborg University Hospital, based on the inclusion criteria and methods found in appendix B

The cell line cultivation and single cell suspension preparation

A single cell suspension useable for flow cytometry was obtained. See appendix C for the single cell suspension procedure including loosening with trypsin, homogenization, extra centrifugation, sterile filtration with 100µM filter and information on HT29 and AMO-1 cell lines

Flow cytometry

Flow cytometry is a technology that can be used to categorize cells according their physical and chemical characteristics. The technology enables cell sorting, cell counting, and biomarker detection when cells pass by in suspension through a tube close to an optical detector. Flow cytometry based on Fluorescence-activated cell sorting (FACS) was performed by experienced flow cytometrists at the Department of Hematology at Aalborg

University Hospital (Alexander Schmitz and Henning Sand Christensen) using the BD FACSAria cell sorter. The equipment allows sorting of e.g. epithelial cells based on their scatter properties and individual antigen expression of combinations of antigens on the cell surface, specifically marked by fluorescence tagged antibodies against these epitopes. The common protocol was based on recommendations from the equipment manufacturer and the preparation of antigen antibody suspension was prepared as described in appendix D. The manual gating strategies were based on the protocol described in (19) and the cytometrist's experience. See figure 2.2 for an example of a standard gating procedure.

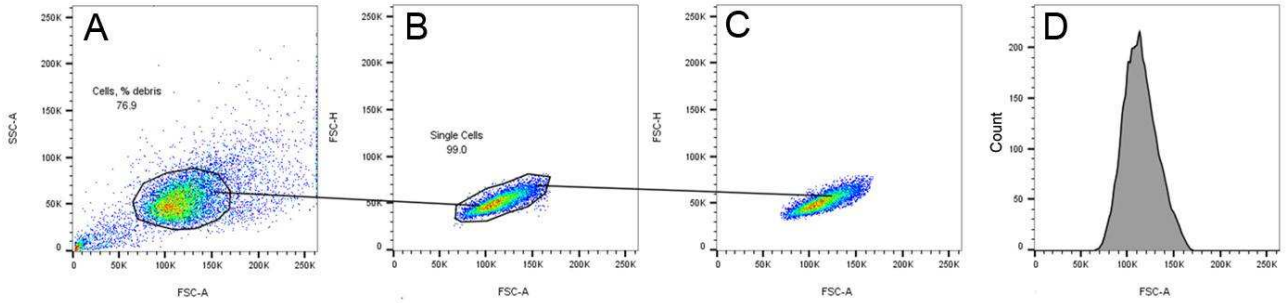


Figure 2.2 shows an example of doublet discrimination on a HT-29 cell line using a FSC-A/FSC-H gating strategy. The Side Scatter (SSC-A) parameter in (A) correlates with cell granularity and is a measure of light reflected at 90 degrees. The Forward Scatter (FSC) is correlated to cell size and is a measure of light scattered less than 10 degrees when a cell passes the laser beam. FSC-A is a measure of area under the pulse curve while FSC-H is the maximum height of the pulse. The FSC-A/SSC-A plot in (A) is used to filter out the debris. Single cells appear in a diagonal on the FSC-A/FSC-H plot (B). By gating the cells on the diagonal it is possible to separate the single cells as shown in (C). A histogram (D) shows the distribution of cell sizes after gating.

Experiment I. Flow cytometry using single stains and various antibodies on normal colonic epithelial tissue

Single cell staining was performed to determine the basic antibody binding efficiency. The following antibodies were used: CD44-PerCP-Cy5.5, CD45–FITS, LGR5-Pe-Vio770, CD324-APS-VIO770, EpCAM-PE, CD24-PE, CD31-FITC, CD166-AF488, EphB2-APC, CD66a-AF488; See appendix E for further information about antibodies used.

A basic gating procedure was performed to discard debris and cell doublets as shown in figure 2.2. The resulting population was divided into three groups according to cell size on the x-axis (FSC-A) and granularity on the y-axis (SSC-A). Group 2 containing the smallest and least complex cells was then discarded as it was not likely that this group could contain the cells of interest. Groups 1 and 3 were assessed on the subject of individual antibody binding abilities by simple eye-balling (see figure 3.1).

Common sources of errors when performing flow cytometry is autofluorescence and unspecific staining. Unspecific staining is usually due to the fact that non-viable cells bind antibodies or that a wrong antibody concentration is used. Therefore it is important to know if non-viable cells are present in the test solution and that the antibody concentration is correct.

Experiment II+III. Viability stains

Viability staining was performed to estimate overall cellularity and viability of the purified primary single cell samples. The acquired events were differentially gated according to their cell size and subsequently analyzed for their binding of 7-AAD which is a marker of non-viable cells. 7-AAD will stain cells with compromised cell membranes indicating that the cells are either apoptotic or dead.

A comparative titration experiment was performed on a HT-29 cell line that was heat treated at 65^0 C for 1 minute according to the instructions from the manufacturer. The cell suspension was mixed with viable HT-29 cells in a 1:1 proportion. FDV-eFlour450 was compared to a 7-AAD viability stain. Unstained gating population cleared from the doublets was used as control.

Experiment IV. Cell line antibody titration experiment

A titration experiment was performed on selected antibodies aimed at reducing unspecific bindings and lowering the cost of the experiment. Gated populations of unstained cells cleared from the small cells, debris and doublets was used as a control population.

Experiment V. Tissue sample characterization of CD45 cell population

The credibility test of certain antibodies was performed on CD45 cell hematopoietic population of mostly leucocytes as it is a well-known cell type characterized with relation to antigen expression. The population was located by gating of populations with different combinations of EpCAM and CD45 expression. A combination of low EpCAM and high CD45 defined a population of leucocytes. Those were examined with regard to expression of CD45 and Lgr5. The cells were extracted during the mortaring process supposedly from peripheral blood.

Experiment VI. Cell line MACS procedure validation experiment

A MACS separation procedure was performed to assess the effective use of the tissue handling procedure. The principle of MACS is based on a magnetically labeled EpCAM stain (Microbeads) that will retain the stained cells within a special magnetic separator while other cells run through. The separate cultures of AMO-1 and HT-29 cells first unstained and stained with EpCAM antibody-linked PE were examined for EpCAM expression. Subsequently, the HT-29 population was mixed with the AMO-1 (1:2) culture and magnetically labeled with EpCAM MicroBeads. The resulting cell suspension was placed in the magnetic field of the MACS separator and the run-through was collected. The tube was then removed from the magnetic field and the positively selected EpCAM cell fraction was collected. Further details are listed in appendix F.

Results

Experiment I. Flow cytometry using single stains and various antibodies on normal colonic epithelial tissue demonstrates possible unspecific staining

Repeated flow cytometries were carried out with different single stains using a panel of antibodies on single cell suspensions from normal colonic epithelial tissue. The aim was to investigate the binding capacity of different antibodies used in previous research to validate the methods for the next steps (18).

A single cell suspension of non-malignant adjacent colon epithelium tissue was obtained using a gentle mortar method. Flow cytometry was performed without staining to assess the side scatter (correlates with cell granularity) and forward scatter (correlates with cell size). The scatter image demonstrated a cell population that could be divided into three groups according to granularity and size. The gating of the groups is shown in figure 3.1. The figure demonstrates the three gated populations. Population 1 named "Cells1" is the greatest with 70.2% of a total cell number and contains the medium sized cells that will be subjected to further analysis. Population two named "Cells 2" contains the smallest cells that will not be further analyzed because the size does not match the expected size of epithelial cells. It makes up 17.0 % of the total cell number and is believed to be haematopoietic cell lines. Population three or "Cells 3" is the group consisting of the largest cells making up 7.5% of the total cell number. Cell population 3 is also further analyzed.

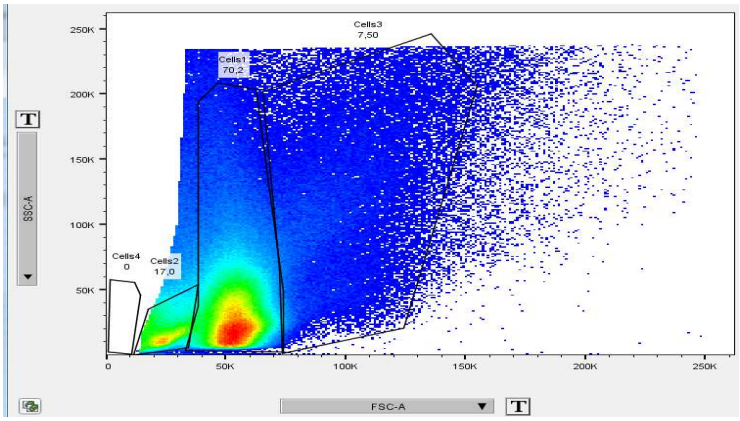


Figure 3.1: The distribution of cell size and granularity within the cell suspension of the colon tissue sample. Data is plotted as cell granularity (side scatter, SSC-A) as a function of cell size (forward scatter, FSC-A). The data has been clustered into three cell populations according to size and granularity by gating. The data outside the gated population represents larger elements, probably doublets.

An example of a single antibody analysis (EpCAM) is shown in figure 3.2. The assessment of individual antibodies in populations "Cells1" and "Cells3" are summarized in table 3.1. The flow cytometry results of all individual stains with interpretation can be found in appendix G. Specificity was evaluated for each individual antibody in single stain analysis. Unstained control was included for determination of unspecific staining. Positive staining was identified for all tested antibodies. The larger "Cells3" population is in general more positive than the smaller "Cells 1". The ratio between "Cells3" and "Cells1" is 8-13 except for CD45 that is used for a different cell type.

EpCAM

Flow cytometry was carried out with EpCAM staining on normal colonic epithelial tissue. The data was divided into the same size/granularity populations as shown in figure 3.1. Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group as shown in figure 3.2 which may indicate autofluorescence. Some fluorescence derived from the EpCAM antibody/fluorochrome was detected outside the unstained cell area.

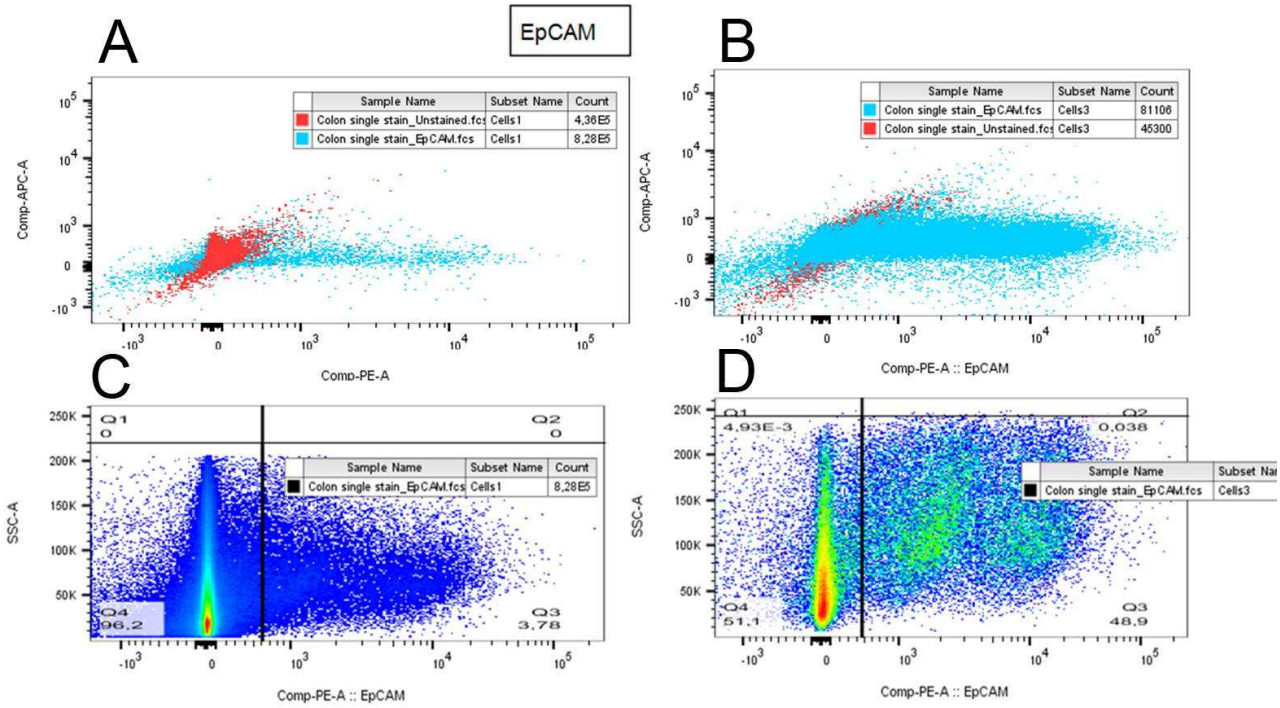


Figure 3.2: The flow cytometry plots of population "Cells1" (A) and "Cells3" (B) using EpCAM staining showing the magnitude of the fluorescence signal compared to an unstained control group. The unstained and stained groups of "Cells1" seem to overlap indicating autofluorescence in the overlapping region. In the "Cells3" group with larger cells there is less overlap between groups. The Y-axis shows fluorescence in other wavelengths than where the antibody is. Scatter plots of EpCAM stained colonic epithelial cells are shown in the bottom of the figure. The scatter images of cell population 1 (C) and 3 (D) demonstrate populations that can be divided into two groups according to granularity and number of events.

The scatter images of cell population 1 and 3 shown in figure 3.2 demonstrate populations that can be divided into two groups according to granularity and number of events. The cell group to the left in (C) is a group of EpCAM negative cells that constitutes 96.2% of the total cell count in population 1. EpCAM negative cells constitute 51.1% of population 3 to the left in (D). The cell group in to the right in (C) is a group of EpCAM positive cells that constitutes 3.8% of population 1. In the right part of (D) it is shown that 48.9% of population 3 was positive. Thus, the two populations exhibit a substantial variation in distribution of the cells in the two groups. Population 1 has a major overweight of EpCAM negative cells while values are more evenly distributed in population 3. There are a substantially higher number of EpCAM positive cells in population 3 as compared to population 1.

Table 3.1 summarizes the results using 10 different antibodies. The scatter plots could either be clustered into 2 or 3 groups with varying degree of overlap between the stained and unstained cells. The percentage of events detected in the most positive population of "Cells1" and "Cells3" are shown in table 3.1. It was expected to see a small ratio using CD45 staining because CD45 is a known marker of haematopoietic cells. The Lgr5 ratio was found to be substantially higher than expected and comparable to other biomarkers. This indicates that the results may be unreliable due to unspecific staining. Because of this it was decided to do a viability test using 7AAD staining.

Table 3.1: The table compares the percent of detected events in the most positive population for "Cells1" and "Cells3" using different antibody staining. The ratio between the percentages detected in the most positive populations is also shown. A higher ratio indicates a larger difference in antibody staining between populations.

Anti-body	Number of groups detected	Cells 1: Percent of events detected in the most positive population	Cells 3: Percent of events detected in the most positive population	Ratio between Cells3 and Cells1	Comment
CD24	3	2.14	18.4	8.6	Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Little fluorescence derived from CD24 was detected outside the unstained cell area.
CD31	2	6.87	67.3	9.8	Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Very little fluorescence derived from CD31 was detected outside the unstained cell area.
CD44	2	5.10	47.7	9.4	Group 1 and 3 demonstrated a lesser overlap with the unstained cell group with a clearer defined delimitation of stained and unstained groups. The stained population is greater displaced to the right on a metric scale and possibly can be divided into two groups
CD45	3	3.61	8.71	2.4	Groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. However the configuration of the stained area differs from the unstained area.
CD66c	2	6.73	74.0	11.0	Both groups 1 and 3 of the stained population demonstrated a certain overlap with the unstained cell group. The configuration of the stained area markedly differs from the unstained area.
CD166	2	9.37	74.4	7.9	Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Only a little fluorescence derived from CD 166 was detected outside the unstained cell area.
CD324	2	5.87	55.8	9.5	Both groups 1 and 3 demonstrated an overlap with the unstained cell group. Some fluorescence derived from CD324 was detected outside the unstained cell area.
EpCAM	2	3.78	48.9	12.9	Both groups 1 and 3 demonstrated an overlap with the unstained cell group. However the configuration of the stained group differes substantially from the configuration of unstained group with a significant amount EpCAM derived fluorescens detected outside the unstained cell area
EphB2	2	5.42	61.7	11.4	Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Very little fluorescence derived from EphB2 antibodies was detected outside the unstained cell area.
Lgr5	2	5.37	50.2	9.3	Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. A little fluorescence derived from Lgr5 antibody was detected outside the unstained cell area.

Due to a high degree of unspecific staining in all immunophenotypic analyses, an unambiguous conclusion was difficult to make. One reason for the unspecific staining could be a high number of non-viable cells in the target population. Another reason could be that the antibody concentration was too high. Further experiments on tissue and cell lines were carried out to investigate the possible influence of those factors.

Experiment II. Tissue sample viability stain demonstrates that non-viable cells stain

Non-viable cells have a tendency of unspecific binding of antibodies in large amounts (20). If the population being investigated contains a lot of dead or dying cells, the results may be biased considerably.

A 7-AAD viability stain was performed on an ungated population from a single cell suspension derived from normal colon tissue epithelium to determine the share of 7-AAD positive cells. The results of the experiment are shown in figure 3.3. In the total ungated population, the number of non-viable cells constituted 21.4% of the total cell count.

The same procedure was performed on 2 gated populations. Population 1 called "Cells" contained 40.5% of the total cell count while population 2 called "Small cells and debris" contained 53.4% of the total cell count. 7-AAD staining of population "Cells" demonstrated 37.8% 7-AAD positive and 62% of 7-AAD negative cells. 7-AAD staining of population "Small cells and debris" demonstrated 4.28% 7AAD positive and 92.2 7-AAD negative cells.

Since the proportion of non-viable cells is 37.8% in the gated "cells" group, it is likely that the results when testing various antibodies, is affected by the unspecific binding to the non-viable cells. When gating the population, the groups containing large cells are relatively more affected by the non-viable cells. This is problematic when the target population we are trying to identify belongs to the group of large cells.

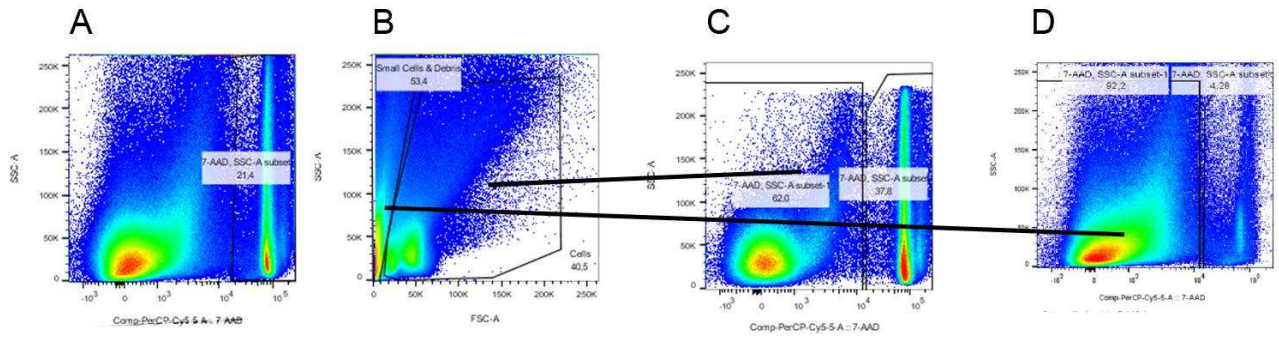


Figure 3.3: The figure shows the result of 7-AAD staining in a population of colon epithelial cells. The unstained population could be divided into 2 groups based on cell size and granularity (B). 7-AAD staining was performed on the entire ungated population (A), the group of large cells (C), and small cells (D).

Experiment III. Cell line viability staining shows that FVD can be used as an alternative to 7-AAD

7-AAD is known to interfere with the fluorescence channel of other antibodies. Therefore it may be difficult to perform multiple stains when 7-AAD is used. It was considered relevant to evaluate both FVD and 7AAD in two concentrations to evaluate if FVD could be used as an alternative. As shown in figure 3.4, the FVD stain demonstrated a very close distribution of values when 1.5µL and 3µL of FVD was used. Both viable and non-viable cells were binding the antibody. Fewer antibodies were bound on viable cells though. Both 7-AAD and FVD stains demonstrated very similar results although the number of non-viable cells in the population was lower than expected as shown in figure 3.4. 7-AAD stain also demonstrates a very close distribution of values when 1.5µL and 3µL of 7-AAD was used. The antibody is bound on viable and non-viable cells with a larger amount bound on non-viable cells. The experiment shows that 1.5µL of FVD can be used as an alternative to 7-AAD and allows a more precise multi-stain procedure.

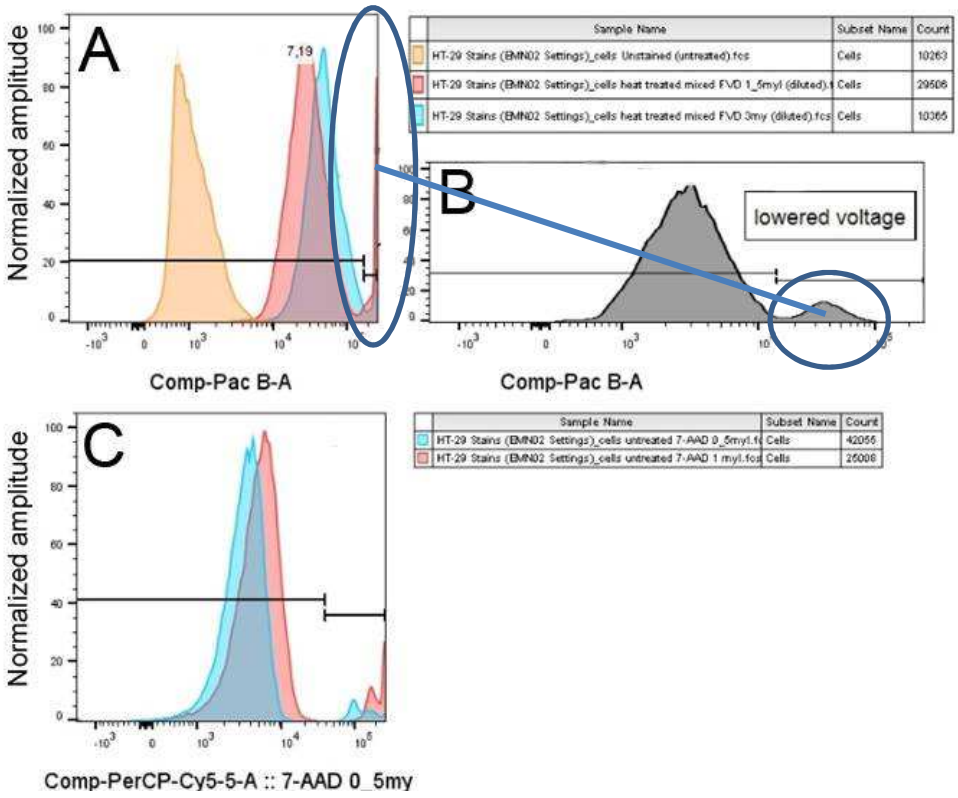


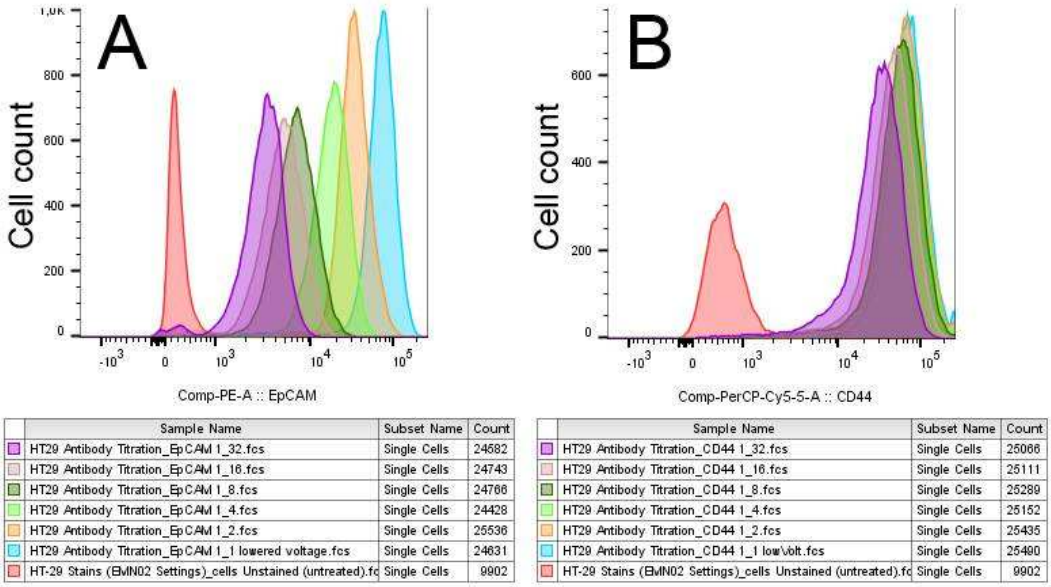
Figure 3.4: A HT-29 cell line sample stained with FVD is shown in (A) in two different concentrations. A two-fold dilution of FVD antibody did not result in a similar shift in positively stained cells even if a decrease was observed. In (B) the voltage was lowered to better visualize the second population assumed to be dead cells. The HT-29 cell line stained with 7-AAD is shown in (C) in two different concentrations.

Experiment IV. Cell line antibody titration experiment shows that the antibody concentration can be optimized

A series of titration experiments were performed to determine the optimal concentrations of the antibodies EpCAM, and CD44. The results of EpCAM titration staining with PE shows a consequent decline in fluorescent signal corresponding to a reduction in the antibody concentrations with no plateau detected and mean values spread between 72232 for 1:1 proportions to 3267 for 1:32 proportions as shown in figure 3.5.

The titration experiment performed on CD44 with PerCP demonstrates a tighter distribution of values, particularly homogeneous at the 1:1, 1:2, 1:4 proportions with mean values 73204, 68814, and 62634 respectively. This distribution corresponds to a plateau configuration for 1:1, 1:2 and 1:4 concentrations.

Based on the findings it only seems possible to reduce the concentration of CD44. It is recommended that similar tests are performed on all biomarkers to determine the optimal concentration.



Sample Name	Subset Name	Count
HT29 Antibody Titration_EpCAM 1_32.fcs	Single Cells	24582
HT29 Antibody Titration_EpCAM 1_16.fcs	Single Cells	24743
HT29 Antibody Titration_EpCAM 1_8.fcs	Single Cells	24766
HT29 Antibody Titration_EpCAM 1_4.fcs	Single Cells	24428
HT29 Antibody Titration_EpCAM 1_2.fcs	Single Cells	25536
HT29 Antibody Titration_EpCAM 1_1 lowered voltage.fcs	Single Cells	24631
HT-29 Stains (BvIN02 Settings)_cells Unstained (untreated).fc	Single Cells	9902

Sample Name	Subset Name	Count
HT29 Antibody Titration_CD44 1_32.fcs	Single Cells	25066
HT29 Antibody Titration_CD44 1_16.fcs	Single Cells	25111
HT29 Antibody Titration_CD44 1_8.fcs	Single Cells	25289
HT29 Antibody Titration_CD44 1_4.fcs	Single Cells	25152
HT29 Antibody Titration_CD44 1_2.fcs	Single Cells	25435
HT29 Antibody Titration_CD44 1_1 low\volt.fcs	Single Cells	25490
HT-29 Stains (BvIN02 Settings)_cells Unstained (untreated).fc	Single Cells	9902

Figure 3.5: Titration of EpCAM (A), and CD44 (B) antibodies on a HT-29 cell line. The distribution of titrated antibodies with proportions 1:1, 1:2, 1:4, 1:8, 1:16, and 1:32 are compared to an unstained sample.

Experiment V. CD45 and LGR5 stain on leukocytes shows unexpected LGR5 binding

The leukocyte population of purified tissue suspension defined by CD45 positivity was investigated for LGR5 binding. The results are shown in figure 3.6. It was expected to see a high CD45 and negative LGR5. However, a flow cytometry image of the antibody staining showed an intense and delimited stain of CD45 and less delimited, but certainly positive stain of LGR5. Since LGR5 binding to leukocytes was unexpected it is necessary to do further experimentation to determine the root cause.

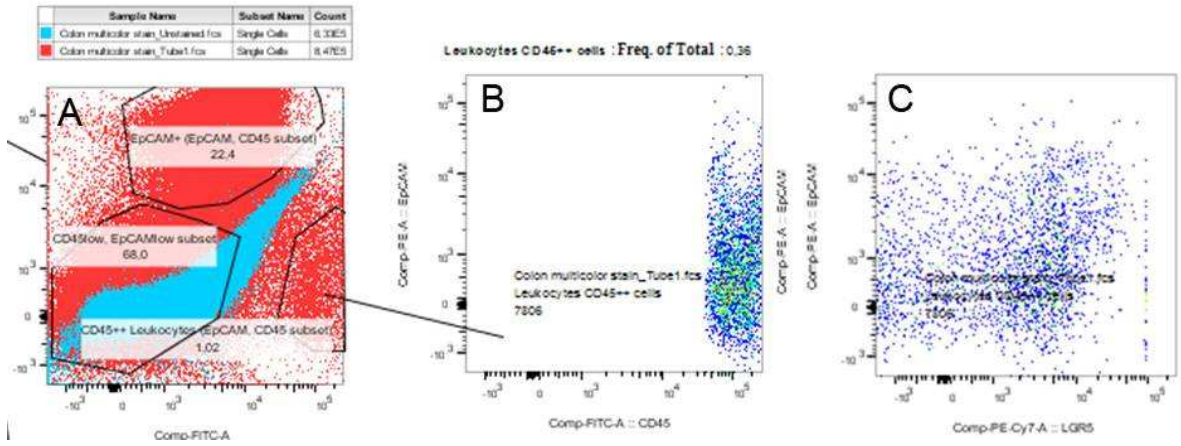


Figure 3.6: A gated population of leukocytes was subjected to a CD45 and LGR5 stain. In (A) the unstained cells are shown in blue color while the cells stained with EpCAM and CD45 are red. As expected, the leukocytes could successfully be stained with CD45 as shown in (B). The LGR5 stain seems unspecific in (C)

Experiment VI. Cell line EpCAM enrichment is possible using the MACS procedure

A magnetic-activated cell sorting (MACS) procedure was tested to assess whether it is suitable for EpCAM enrichment of single cell suspensions from colonic epithelial tissue. MACS separation analysis was initiated by a verification of absence or presence of EpCAM positive cells in AMO-1 and HT-29 cell cultures. AMO-1 culture

demonstrated no EpCAM positive cells both in unstained and stained assays. The HT-29 culture demonstrated 0.88% of EpCAM positive cells in the unstained culture and 99.1% of EpCAM positive cells in the stained culture (see figure 3.7). After the two cultures were mixed prior to enrichment 99.3% of the total cell count was EpCAM negative while a fraction of 0.73% was EpCAM positive.

After the MACS procedure the mixture of the culture demonstrated almost a 100 fold enrichment of EpCAM positive cells (from 0.7% to 68.3%). Therefore, it seems likely that the MACS procedure using EPCAM is suitable for future experiments.

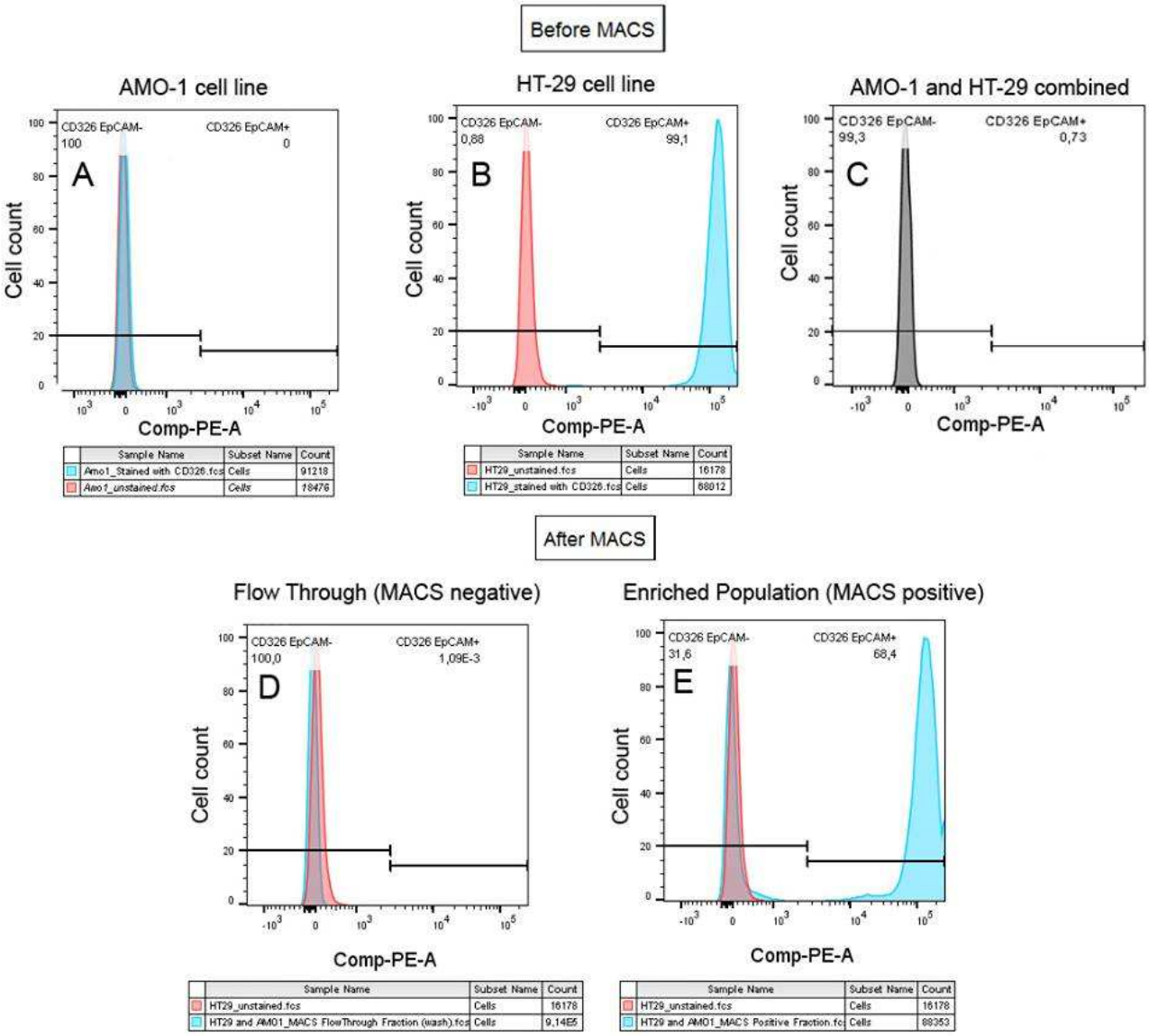


Figure 3.7: EpCAM stain of AMO-1 (A) and HT-29 (B). Alone, the AMO-1 did not stain for EpCAM while the HT-29 stained for EpCam. When the cell cultures are combined (C), the proportion of EpCAM negative and positive cells are 99.3% and 0.7% respectively. EpCAM staining after MACS shows an EpCAM negative cell population in the flow-through (D) and substantially enriched EpCAM positive population (E).

Discussion

The overall hypothesis of this project is that application of differentiation specific gene expression signatures obtained from normal colon epithelial tissue can add prognostic value by identifying whether cancer cells originate from the apical (CMS3 equivalent) or basal (CMS2 equivalent) part of the crypt. The hypothesis is tested in a three- step approach where the first step (step 1) was to characterize and assess the degree of differentiation of normal colon epithelial cell populations using multi-parameter flow cytometry (FCM). The successive steps of the hypothesis testing includes fluorescence assisted cell sorting (FACS) of normal colon epithelial cells based on the degree of differentiation and gene expression profiling of these subsets (step 2) and development of a bioinformatics algorithms of colon-epithelial-associated-gene-signatures that can be used for classification and prognostic evaluation in clinical CRC cohorts (step 3).

The aim of the present study was focused solely on step 1 where single cell suspensions of normal colonic epithelial tissue was stained with fluorescently labeled monoclonal antibodies directed towards surface expressed specific antigens characteristic of undifferentiated ISC originating from the crypt and fully differentiated epithelial cells from the apical region of the intestinal epithelial tissue. Pilot studies were performed and several potential technical problems were identified and addressed through series of experiments on normal colon tissue and human colorectal adenocarcinoma cell line HT-29. The suggested refinement of the method in the cellular identification algorithm can potentially be used in future pursue of step 2 and 3.

Initially, a single stain experiment (Experiment I) was performed to evaluate the overall binding capacity of selected antibodies. Based on size and granularity of analyzed colonic epithelial cells, two relevant cell populations were selected for further analysis; e.g. the larger cells in the “Cells 3” population and the smaller cells in the “Cells1” population. In general, populations “Cells3” were observed to bind higher fractions of antibodies when compared to “Cells1” (Table 3.2)

An exception was when analyzing staining with an antibody directed against CD45. CD45 is a marker for the hematopoietic cell line, mostly leucocytes that are characterized by a smaller cell size comparing to larger epithelial cells. CD45 is expected not to stain epithelial cells. This explains the low ratio when the CD45 stain is used.

In experiment I, the results may be unreliable because of the very similar ratios between the number of events detected in the “Cells1” and “Cells3” populations when comparing the various antibodies. Based on previous research (13)(21), the expression of CD44 and LGR5 is expected to be different with a substantially higher expression of CD44 when compared to Lgr5. This experiment found, however, that the percentage of cells in the “Cells3” group when quantifying CD44 antibody is 47.7% while the same number for Lgr5 is 50.2%. Furthermore, most tested antibodies demonstrate a very similar and high numeric value in the “Cells3” group. This suggests even levels of antigen CD44 and Lgr5 expression for most antibodies tested which is a questionable result. A technical explanation for the unexpected findings in experiment I could be unspecific staining. Different sources can cause unspecific staining such as presence of non-viable cells, choice of antibody + flourochrome, and concentration of the applied antibodies. The large amount of fluorescence emitted by non-viable cells was further investigated in a series of viability stain experiments, both on tissue samples and on a HT-29 cell line (Experiment II+III).

In experiment II performed on a tissue based suspension it was demonstrated that the viability stain of the ungated population contained 21.4% of 7-AAD positive and thus non-viable cells. This is a proportion large enough to affect the results. Therefore, non-viable cells should be eliminated from the target population in future experiments. Further analysis was performed on the same population in experiment II due to the fact that several small viable cells, including erythrocytes and debris are also 7-AAD negative. A gating of the population was performed using a 7-AAD stain. It was expected that the population of “Small cells/debris” fraction should

stain less with 7-AAD. That was confirmed implying smaller amount of nucleated apoptotic cells in this group. The population "Cells" demonstrated a substantially higher fraction of non-viable cells making it a main target for identification of epithelial cells. However, since the proportion of non-viable cells is 37.8% in the gated "Cells" group, it is likely that the results when testing various antibodies, is affected by the unspecific binding of the non-viable cells. When gating the population, the groups containing large cells are relatively more affected by the non-viable cells. This is problematic when we consider the target population belongs to the group of large cells. Overall, unspecific staining of non-viable cells were documented in both "small cell/debris" and "ungated population expected to contain epithelial cells of interest supporting that future analysis should take viability stains into consideration to avoid unspecific staining.

In experiment III, the quality of the viability stain was additionally investigated on a HT-29 cell line. One of the issues addressed here was the optimization of the concentration necessary for an optimal viability stain. If used in larger concentrations, the staining could give an unspecific staining and a 50% reduction of the concentration of antibody was therefore tested together with an alternative FDV viability staining. The use of FDV would be beneficial as FVD stain uses another fluorescence channel than 7-AAD and is better suited for the multistain procedures of the antibodies chosen in this project.
Both FVD and 7-AAD viability stains showed a very close distribution when using 1.5 and 3µL of antibody (figure 3.5). That supports the possibility of using a smaller amount of antibody in the future tissue experiments. FVD stain showed the same result as 7-AAD and is therefore recommended in future experiments.

Another source of unspecific staining could be caused by suboptimal antibody concentration. As mentioned earlier the higher concentration of the antibody can cause unspecific staining. The most industrial antibody kits are accompanied by an instruction describing an optimal antibody concentration. However, in some cases the interplay between certain antigens and antibodies leads to modified optimal concentrations (22)

In experiment IV, titration optimization demonstrated that the antibody concentration can't be lowered on PE staining of EpCAM or CD24 as the reduction of antibody concentration correlated with reduction of staining intensity. The titration experiment performed on CD44 with CompPerCP-Cyp5-5-a demonstrates binding intensities independent of antibody concentration covering the range 1:1, 1:2, 1:4, and 1:8 supporting that antibody concentrations can be reduced without losing stoichiometric saturation on the cell surface. This indicates that it is possible to reduce the antibody volume needed for a successful CD44 detection and for decreasing the level of unspecific antibody binding.

In experiment V, the Lgr5 stain demonstrated questionable results in the "Cells3" population. This is concluded on basis of the high detected concentrations when compared to the other markers. Lgr5 was believed to be low in the tissue (13). Therefore, the finding of high LGR5 concentrations in this study raises a concern on credibility of the Lgr5 stain. The concern became strengthened by a characterization of CD45 cell population where LGR5 is again found positive despite the assumption that CD45 cells don't express Lgr5. Lgr5 is one of the central antigens when talking about CRC. Further validation of the antibody is absolutely necessary and should be carried out in the future research.

Other approaches to optimize ISC identification and characterization include MACS. In experiment VI, a validation and optimization of the Lgr5 stain was investigated since it would allow a better characterization of cells from the base of the crypt with regard to expression of other more frequent antigens. That could be done by isolation of a group of Lgr5 and EPCAM positive cells by MACS and perform further studies on that population. This culture enrichment procedure is an example of a principle described in the earlier studies on HT-29 cell line (23). This study doesn't provide the exact information on how the culture was enriched, but provides evidence of the usability of the procedure in context of Lgr5 bearing cell population studies.
In this project, a principle enrichment by MACS validation experiment was successfully performed on a HT-29 cell line (experiment VI), but could not be done on tissue sample partially because of compromised Lgr5 staining

and partially because of low cell yield during the tissue preparation procedure. The first concern could be addressed by more extensive research on antibody optimization both on primary tissue and cell lines. The second concern might be solved by optimization of the tissue procurement procedure by collection of a larger tissue sample.

This study was performed as part of the first step in the process of identifying and characterizing cells from the basal and apical intestinal crypts. The subsequent steps will include a global gene expression analysis and comparison to global gene expression datasets on CRC. The identified gene signatures could potentially create a prognostic tool for stratification of CRC cases that can prove valuable in clinical practice. The most recent classification was provided in "Consensus molecular subtypes of colorectal cancer" (9). This study proposed a classification of CRC based on gene expression signatures derived from studies of cancerous tissue (9). The present research aims to provide a classification based on gene analysis with a starting point using normal tissue – a concept that has documented advantages in disease classification and prognostic assessment in B-cell malignancies (18). The advantage of this approach is that it uses the normal and reproducible stable gene signatures of differentiation epithelial cells as a starting point for comparison to the cancerous cases. Molecular subtyping based solely on cancerous tissue is dependent on tumor biopsy cellular composition enabling significant uncertainty to the classification strategy.

Some research has investigated the feasibility of implementing the consensus GEP based CMS1,2,3,4, molecular subtype system of CRC to clinical practice within the field of pathology (10). According to this study the degree of affiliation to the CMS2 group can be assessed by quantifying cell proliferation. This can be done by different means such as flow cytometry and immunohistochemical Ki67 staining. Ki67 is a relatively cheap alternative with a robust staining of proliferating cells that can be performed automatically. Unfortunately, there is a lack of robust markers when distinguishing CMS3 cases. Similar approaches can be used for distinguishing the cancer cases derived from the cells from the basal and apical intestinal crypts. If further research on the topic demonstrates a promising method to characterize CRC tumors, it may prove to be clinically very relevant.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.

2. Dccg. Landsdækkende database for kræft i tyk - og endetarm (DCCG) National årsrapport 2014. Dccg. 2016;(december):1–182.

3. Harling H. Landsdækkende database for kræft i tyktarm og endetarm Årsrapport 2009. Dccg. 2004;1–57.

4. Tnm-klassifikation K, Tnm-klassifikation K, Tnm-klassifikation MP, Controls IC. Dccg.Dk Nationale Retningslinier for Diagnostik Og Behandling Af Kolorektal Cancer. 2017;1–17. Available from: http://dccg.dk/retningslinjer/2017/2017_Tumorklassifikation_TNM_ver 1.1.pdf

5. The Danish Colorectal Cancer Group. Histological classification. 2017;1–2.

6. Harling H. Landsdækkende database for kræft i tyktarm og endetarm Årsrapport 2009. Dccg. 2010;

7. Prognosticering PT, Kvalitetsvurdering T. Mikroskopi af kolorektalcancer resektater. 2017;(1):1–14.

8. Halvorsen TB, Seim E. Degree of differentiation in colorectal adenocarcinomas: a multivariate analysis of the influence on survival. J Clin Pathol. 1988;41(5):532–7.

9. Guinney J, Dienstmann R, Wang X, De Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350–6.

10. Roseweir AK, McMillan DC, Horgan PG, Edwards J. Colorectal cancer subtypes: Translation to routine clinical pathology. Cancer Treat Rev [Internet]. 2017;57:1–7. Available from: http://dx.doi.org/10.1016/j.ctrv.2017.04.006

11. Han S, Yang W, Zong S, Li H, Liu S, Li W, et al. Clinicopathological, prognostic and predictive value of CD166 expression in colorectal cancer: a meta-analysis. Oncotarget [Internet]. 2017;8(38):64373–84. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=28512258

12. Lemischka KM and I. Linked references are available on JSTOR for this article : 2006;25(2):75–8.

13. Cui S, Chang P-Y. Current understanding concerning intestinal stem cells. World J Gastroenterol [Internet]. 2016;22(31):7099. Available from: http://www.wjgnet.com/1007-9327/full/v22/i31/7099.htm

14. Basu S, Haase G, Ben-ze A. Wnt signaling in cancer stem cells and colon cancer metastasis [version 1 ; referees : 3 approved] Referee Status : 2016;5(0):1–10.

15. Kumar KK, Burgess AW, Gulbis JM. Structure and function of LGR5: An enigmatic G-protein coupled receptor marking stem cells. Protein Sci. 2014;23(5):551–65.

16. Won-Tae K, Chun JR. Cancer stem cell surface markers on normal stem cells. BMB Rep. 2017;50(BMB-17-039):285–98.

17. Hatano Y, Fukuda S, Hisamatsu K, Hirata A, Hara A, Tomita H. Multifaceted interpretation of colon cancer stem cells. Int J Mol Sci. 2017;18(7).

18. Selskab for Medicinsk Studenterforskning. 2017.

19. Magness ST, Puthoff BJ, Crissey MA, Dunn J, Henning SJ, Houchen C, et al. A multicenter study to standardize reporting and analyses of fluorescence-activated cell-sorted murine intestinal epithelial cells. AJP Gastrointest Liver Physiol [Internet]. 2013;305(8):G542–51. Available from: http://ajpgi.physiology.org/cgi/doi/10.1152/ajpgi.00481.2012

20. Kuonen F, Touvrey C, Laurent J, Ruegg C. Fc block treatment, dead cells exclusion, and cell aggregates discrimination concur to prevent phenotypical artifacts in the analysis of subpopulations of tumor-infiltrating CD11b+myelomonocytic cells. Cytom Part A. 2010;77(11):1082–90.

21. Senbanjo LT, Chellaiah MA. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front Cell Dev Biol [Internet]. 2017;5(March). Available from: http://journal.frontiersin.org/article/10.3389/fcell.2017.00018/full

22. Hulspas R, O'Gorman MRG, Wood BL, Gratama JW, Robert Sutherland D. Considerations for the control of background fluorescence in clinical flow cytometry. Cytom Part B - Clin Cytom. 2009;76(6):355–64.

23. Chen X, Wei B, Han X, Zheng Z, Huang J, Liu J, et al. LGR5 is required for the maintenance of spheroid-derived colon cancer stem cells. Int J Mol Med. 2014;34(1):35–42.

24. Schnell U, Cirulli V, Giepmans BNG. EpCAM: Structure and function in health and disease. Biochim Biophys Acta - Biomembr [Internet]. 2013;1828(8):1989–2001. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0005273613001338

25. Patriarca C, Macchi RM, Marschner AK, Mellstedt H. Epithelial cell adhesion molecule expression (CD326) in cancer: A short review. Cancer Treat Rev [Internet]. 2012;38(1):68–75. Available from: http://dx.doi.org/10.1016/j.ctrv.2011.04.002

26. Chai XB, Song RF, Xu F. Expression changes in epithelial cell adhesion molecule during colorectal cancer tumorigenesis. 2015;14(3):7624–9.

27. Han S, Zong S, Shi Q, Li H, Liu S, Yang W, et al. Is Ep-CAM Expression a Diagnostic and Prognostic Biomarker for Colorectal Cancer? A Systematic Meta-Analysis. EBioMedicine [Internet]. 2017;20:61–9. Available from: http://dx.doi.org/10.1016/j.ebiom.2017.05.025

28. Thapa R, Wilson GD. The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. Stem Cells Int. 2016;2016(Figure 1).

29. Wijnhoven BP, Dinjens WN, Pignatelli M. E-cadherin-catenin cell-cell adhesion complex and human cancer. Br J Surg. 2000;87(8):992–1005.

30. Morimoto T, Mitomi H, Saito T, Takahashi M, Murakami T, Sakamoto N, et al. Distinct profile of HIF1??, PTCH, EphB2, or DNA repair protein expression and BRAF mutation in colorectal serrated adenoma. J Gastroenterol Hepatol. 2014;29(6):1192–9.

31. Batlle E, Bacani J, Begthel H, Jonkeer S, Gregorieff A, Van De Born M, et al. EphB receptor activity suppresses colorectal cancer progression. Nature. 2005;435(7045):1126–30.

32. Alazzouzi H, Davalos V, Kokko A, Domingo E, Woerner SM, Wilson AJ, et al. Mechanisms of inactivation of the receptor tyrosine kinase EPHB2 in colorectal tumors. Cancer Res. 2005;65(22):10170–3.

33. Tachezy M, Zander H, Gebauer F, Marx A, Kaifi JT, Izbicki JR, et al. Activated leukocyte cell adhesion molecule (CD166) - Its prognostic power for colorectal cancer patients. J Surg Res [Internet]. 2012;177(1):e15–20. Available from: http://dx.doi.org/10.1016/j.jss.2012.02.013

34. Ribeiro KB, Da Silva Zanetti J, Ribeiro-Silva A, Rapatoni L, De Oliveira HF, Da Cunha Tirapelli DP, et al. KRAS mutation associated with CD44/CD166 immunoexpression as predictors of worse outcome in metastatic colon cancer. Cancer Biomarkers. 2016;16(4):513–21.

35. Wang W, Zhang G, Yang J, Gu H, Ding L, Yu H, et al. Digital gene expression profiling analysis of DNA repair pathways in colon cancer stem population of HT29 cells. Acta Biochim Biophys Sin (Shanghai). 2017;49(1):90–100.

36. Atashpour S, Fouladdel S, Movahhed TK, Barzegar E, Ghahremani MH, Ostad SN, et al. Quercetin induces cell cycle arrest and apoptosis in CD133(+) cancer stem cells of human colorectal HT29 cancer cell line and enhances anticancer effects of doxorubicin. Iran J Basic Med Sci. 2015;18(7):635–43.

Appendix A

EpCAM

The epithelial cell adhesion molecule EpCAM is a very recognized marker belonging to the epithelial cell population (24), which covers epithelial tissues including neoplasms derived from those(25). It is absent in connective tissues and squamous epithelium and is also known as CD326 (26). EpCAM is a cell surface glycoprotein with a weight of approximately 40 kDa. It plays a morphoregulatory role and is involved in proliferation, migration, adhesion, differentiation, and cell signaling in normal epithelia, stem and progenitor cells. (24), (25)

In normal adult tissue, EpCAM is localized in basolateral cell membranes in simple, pseudo-stratified, and transitional epithelia in most organs and glands. The level of expression varies between tissues with highest levels in the colon and generally in proliferating cells, while lower levels of expression are detected in differentiated cells(24). EpCAM gradient is observed in the intestinal epithelium through crypt to villus axis with highest expression found at the base. EpCAM presence is also found in in stem and progenitor cells of other organs such as hepatocytes during embryonic development or adult regeneration and fetal epithelia. Thus EpCAM is defined as surface marker of pluripotent human embryonic stem cells (24)

EpCAM was discovered as one of the first cancer markers being highly expressed in epithelial cancers, but is also now known to be present in lower levels I normal simple epithelia (24). Wide research in the field now provides extensive data on EpCAM expression in various normal and cancerous tissues. Thus, EpCAM expression is found positive in normal colon tissue and enhanced in both colon adenomas and CRC (25), (27), (26), (26) shows the positive expression rate of EpCAM in CRC at 97% while it is lower in normal tissues and normal blank controls with 14% and 12% respectively. The frequency and levels of expression of EpCAM in normal vs adenomas and CRC tissues varies across different studies (27) That makes EpCAM likely to be able to drive tumor progression in cancer cells(24), possibly through Wnt signaling pathways discussed earlier (26) and through epithelial to mesenchymal transition (27)

Due to the tumor specific overexpression, EpCAM has been a target for various immunotherapeutic approaches (26), such as treatment with monoclonal antibodies and vaccination approaches (24)(25)

CD44

CD44 is a surface protein also referred to as P-glycoprotein, largely distributed through the body (21) and encoded by CD44 gene, a highly complex gene consisting of 19 exons large share of which are variant exons (CD44v) (28)CD44 variants have been found in both normal and cancerous tissue of different differentiation (28) The standard iso-form is found in a large variety of tissues while variant forms are restricted to selection of epithelial cells (21). The normal function of CD44 is maintainance of hyaluronic metabolism, activation of lymphocytes, and release of cytokines (21). Several variant forms have shown to be involved in malignant tumors and their metastasis with different role and resulting degree of malignancy (21), (28). Multiple signaling pathways might be involved, such as pathways activated by tyrosine kinases and Ras family, leading to cellular proliferation, invasion and chemoresistance (28). CD44v3 is one of the variants associated with poor prognosis in CRC being able to initiate invasion and resist apoptosis (28). CD44v6 is another variant in CRC associated with higher metastatic activity and correlated with lower disease-free survival. It has been detected on populations of CSCs, making the protein a potential biomarker for the identification of those (28). Also CD8-10 have shown to play a role in metastasis and be able to serve as an independent prognostic factor(28). Targeted therapy against CD44 is now being investigated (28)

CD-324

CD324 also named CDH1 is a marker of e-cadherin, important cell to cell adhesion molecule in epithelial tissues. It helps with arranging epithelial cells in sheets and loss of those can indicate cancerous epithelial to mesenchymal transformation which can promote metastatic processes (29). Immunohistochemical studies indicate that differentiated and less invasive tumors demonstrate strong staining for e-cadherin while poorly differentiated tumors that show a strong invasive behavior demonstrate a reduced stain (29). It is hypothesized that the loss of cadherins is both involved in early and late cancerogenic events, as the cancer cells that have lost

their adhesiveness might be able to detach from their primarily site and invade through tissues facilitating a local spread or creating a basis for metastatic processes in the distant organs (29).

EphB2

EphB2 is a member of the receptor tyrosine kinase family and is responsible for several signaling pathways related to cell growth, migration, and sustaining crypt-villus axis in the intestinal crypt (30). EphB2 is found in the epithelial cells of normal crypts (31) while reduced levels of EphB2 have been identified in CRC (30). The reduction in EphB expression correlates strongly with the degree of malignancy. (31). Though significantly increased expression of EphB2 detected in premalignant colorectal tissues with activated Wnt pathway (31). Thus EphB2 is assumed to be a direct transcriptional target of the TCF/h-catenin pathway in premalignant lesions which is now viewed as a hallmark of colorectal cancer (32); Likewise the low expression of EphB2 in CRC expresses the importance in colorectal tumor progression (32).

CD166

CD166 or ALCAM, activated leukocyte cell adhesion molecule is a glycoprotein with an important role in different tissues during embryogenesis. It is thought to be involved in tumorogenesis (33) and now is viewed as an important CSC marker in CRC (11), (34) and (33) as its expression is correlated with cancerous processes of various types inclusive CRC (11). CD166 is found in an adult mucosa, but is significantly increased in colorectal tumor cells compared to both colonic adenomas and normal colonic mucosa (11). It has been investigated in relation to its prognostic value for CRC and studied as a potential target for therapy, although data is inconsistent (33)(11)

CD133

CD133 or prominin-1 is a multifunctional transmembrane protein (Diff, Hatano 2017) found on stem like cells and carcinomas of different tissues colon inclusive (CD133, Al alsary 2017),(CD133, Schmohl, 2016). The precise physiological function of the protein is unknown even though it seems to be associated with Wnt pathway and thus differentiation (CD133, Schmohl, 2016) It might also be involved in cell to cell and cell to matrix interactions, found in many normal epithelial cells and in stem cells of adipose and intestinal tissues, (EpCAM Won Toe, 2017) In cancer tissues it is thought to be affecting stemness, tumorigenesis, chemo-/radio-resistance, metabolism, autophagy, and apoptosis (Diff, Hatano 2017). A correlation has been detected between the overexpression of the protein and poor patient survival and several drugs selectively targeting CD133 has been developed (CD133, Schmohl, 2016).

CD24

CD24 or Heat stable antigen is B cell-related short glycoprotein (Diff, Hatano 2017) located on external membrane of the cell (CD24, Pan, 2016) (CD24, Sahlberg 2014). It promotes cell cycle via increased proliferation and inhibited apoptosis (CD24, Pan, 2016). Expression is associated with different types of cancer such as gastric, pancreatic, and breast cancer (Diff, Hatano 2017), but it is also seen on adult intestinal stem cells (EpCAM Won Toe, 2017). It is rarely found in normal tissues apart from B cell precursors, neutrophils, neuronal cells, and certain epithelial cells (?Diff, Hatano 2017). The cells are characterized by high chemotherapy resistance, self-renewal ability and tumorigenic, potentially making them a likely CSC candidate (CD24, Pan, 2016).

CD66a

CD66a also known as CEACAM1 is an adhesion molecule of the carcinoembryonic family, CEACAM, and immunoglobulin superfamily (CD66a, Heine, 2011). CEACAM1 is found on various epithelial cells and different leukocytes with 11 different splice variants on humans detected (CD66a, Fiori, 2012). CEACAM1 is viewed as a tumor suppressor, because the restauration of its function in some tumor cell lines lead to reduced oncogenicity.

CD45

CD45 is a receptor-type protein tyrosine phosphatase and is commonly expressed on all nucleated hematopoietic cells. (CD45, Rhee, 2012)

CD31

Vascular endothelium marker that helps to identify blood vessels

HT29

HT29 is a human colorectal adenocarcinoma cell line. Extensive research has not been done on isolating and mapping the colon cancer specific antigen panel, however, HT29 is known for expressing among others CD44 (HT29, Wang 2016) and CD133 (HT29, Atashpour 2015). A successful attempt of isolating cancer stem cells in a HT29 sphere forming cell culture has been done based on their expression when combining the two surface markers with ALDH1+ (HT29, Fan 2011). (HT29, Chen, 2014) isolated Lgr5 positive cells in a HT29 cancer stem cell enriched spheroid culture and found that downregulation of this gene also lead to decreased expression of CD133 and CD44.

Appendix B

Normal colon tissue samples were acquired from resections performed at Aalborg University Hospital, Department of Surgery based on the following inclusion criteria: scheduled for left-sided hemicolectomy or sigmoidectomy, absence of medical history of chemotherapy or radiation therapy treatment within the previous year, no predisposition of either Lynch syndrome or familial adenomatous polyposis (FAP), and no inflammatory bowel diseases such as Crohn's disease or ulcerative colitis.

Written and oral information about the project was provided to patients that met the inclusion criteria during the preoperative hospital appointments. Patients had a 24 hour time period for reflection before final consent had to be provided. Patients were orally reassured that the decision would not affect the surgical treatment.

Normal colon tissue was sampled from the distal or proximal resection margin of the colon specimen as far as possible from the cancerous area and at least 10 cm proximal or distal from a tumor, to ensure an optimal later pathological assessment. The sample collection was performed by a surgeon and was a circumferential sample with a width of 1.5 cm.

In connection with surgical resection, the specimens were collected in the operating room and transported on ice to the laboratory followed by immediate tissue preparation.

Based on previous preparation experiences at our lab by Majken Vestergaard, a mortar method was selected for tissue preparation. The protocol is described in appendix A.

After the described tissue preparations were completed, experiments on tissue samples were carried out using flow cytometry analysis on each cell suspension. See Appendix B for the common protocol for flow cytometry on colon cells

Mortar method for tissue preparation

1. Ambient fat layer is cut away.
2. The tissue is rinsed into NaCl removing the faeces residues
3. The circumferential colon sample is cut across leaving a long strip of tissue
4. The mucosa and submucosa layers are dissected
5. The remaining tissue sample is weighed
6. The tissue sample is cut into pieces 1*2 each and placed into the mortar together with approximately 6 ml PBS
7. The tissue sample was gently mortared for a couple of minutes
8. The cell clumps is placed in a 50 ml glass while the cell suspension is placed in a 50 ml tube for itself.
9. 50 ml tubes with cell lumps are centrifuged at 500 g for 5 min at 4 degrees.
10. The supernatant is collected and fed into the tube with the cell suspension.
11. The cell lumps are discarded
12. The cell suspension is filtered- first through 100 µm, then 40 µm filter.
13. The cell suspension is centrifuged at 500 g. For 5 min at 4 degrees.
14. The supernatant is removed and pellet is re-dissolved in 0.5 ml PBS (dilution x 10)
15. The cell count is estimated by… method

Appendix C

HT29 is a human colorectal adenocarcinoma cell line. Extensive research has not been done on isolating and mapping the colon cancer specific antigen panel, however, HT29 is known for expressing among others CD44 (35) and CD133 (36). (23) isolated Lgr5 positive cells in a HT29 cancer stem cell enriched spheroid culture and found that downregulation of this gene also lead to decreased expression of CD133 and CD44.

AMO1 is a CD4-positive plasmacytoma cell line that was used for dilution in the MACS procedure experiment on cell lines.

Single cell suspention procedure on CACO2 and HT-29

1. Soak media from, add 10ml PBS (1x) and rinse the surface of cells to remove media from the medium and suction
2. Add 5ml trypsin and incubate 3 minutes into CO2 incubator - unblock the cells from the culture bottle using cell scrapers to loosen the cells
3. Add 10ml media and transfer to centrifuge tubes - make sure the suspension is as homogeneous as possible, spin 1200 rpm for 5 minutes and re-release in new media and count using hemacytometer.

Appendix D

The common protocol for flow cytometry on colon cells

Protocol performed on ice.

1. Add Staining Buffer with BSA (SB) (4°C) to sample tube and centrifuge 5min at 540G 4°C.
2. Remove supernatant and resuspend in 200µl SB. Mix gently.
3. Add antibodies to tubes appropriately and mix gently. Incubation for 30min at 4°C.
4. Add 2ml SB to wash cells. Centrifuge for 5min at 540G 4°C.
5. Remove supernatant and resuspend in 500 - 800µl SB depending on cell concentration.
6. Samples stored in refrigerator until analysis. FACSAria II sample chamber set to 100rpm and 4°C.

Appendix E

An overview of antibodies used in initial single stain procedure aimed preliminary identification of antibodies potentially useable for delimitation of the target population of ISCs.

6/10-17	Antibody	Fluorochrome	Volume	Lot#	Producer
	CD44	PerCP-Cy5.5	5µl	5170520	BD
	CD45	FITC	20µl	5223685	BD
	LGR5	PE-Vio770	5µl	5151214054	Miltenyi
	CD324	APC-Vio770	5µl	5151214055	Miltenyi
	EpCAM	PE	20µl	5222664	BD
	CD24	PE	20µl	88658	BD
	CD31	FITC	20µl	5090933	BD
	CD166	AF488	10µl	1207	AbD
	EphB2	APC	5µl	5247606	BD
	CD66a	AF488	5µl	B191974	Biolegend
	7-AAD		20µl	2213673	BD

1. Single stain tubes.

Appendix F

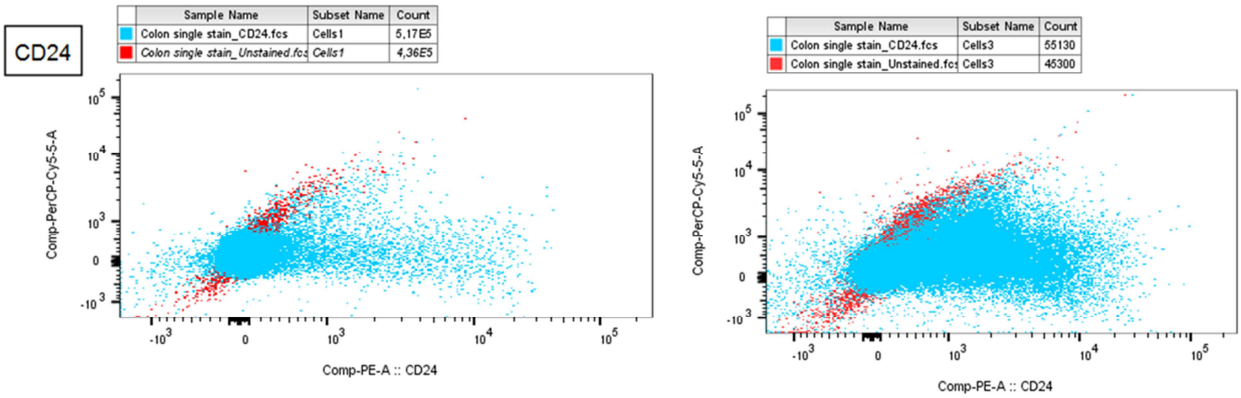
HT29 single cels in a suspension of AMO-1.

4. Tubes included are Unstained, 1%, 0,5%, 0,1% and 0,01%.
5. 5Mio celler HT29 for 10Mio celler AMO-1
6. Centrifuge cells 10min at 800G.
7. Remove supernatant and resuspend HT29 in 2ml SB (Staining buffer) and AMO-1 in 1ml SB.
8. Further diluting some of HT29 (1:20, 8µl HT29 + 152µl SB) for Tube 0,1% and 0,01%.
9. Stain and incubate samples for for 30min room temp in darkness.
10. Wash samples with 2ml SB.
11. Centrifuge 5min at 540G.
12. Remove supernatant and resuspend cells in 200µl SB.

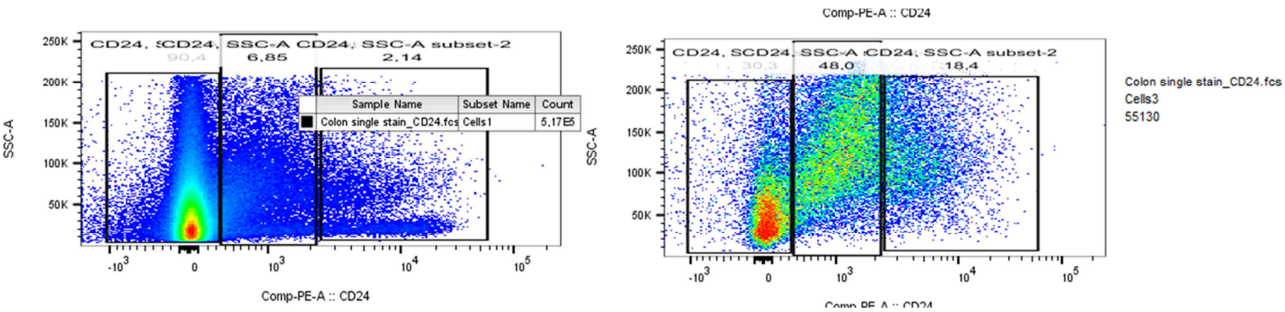
Appendix G

CD24

Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Some fluorescence derived from CD24 antigen was detected outside the unstained cell area.

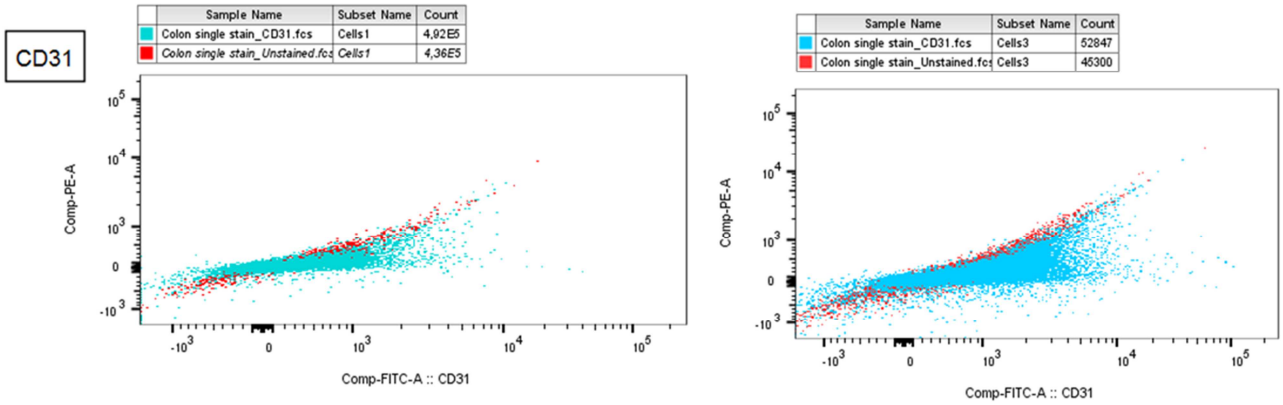


The scatter images of cell population 1 and 3 demonstrate populations that can be divided into three groups according to granularity and number of events. The cell group to the left is a group of CD24 negative cells that constitutes 90.4% of the total cell count in population 1. CD24 negative cells constitute 30.3% of population 3. The group in the middle has an intermediate expression of CD24 antigen and constitutes 6.85% and 48.0% of population 1 and 3 respectively. The cell group to the right is a group of CD24 positive cells that constitutes 2.14% of population 1 and 18.4 % of population 3. Thus, the two populations exhibit a substantial variation in distribution of the cells in the three groups. Population 1 has a major overweight of CD24 negative cells while values are more evenly distributed in population 3 where the largest number of cells can be placed in the intermediate category. There is a substantially higher number of CD24 positive cells in population 3 as compared to population 1.

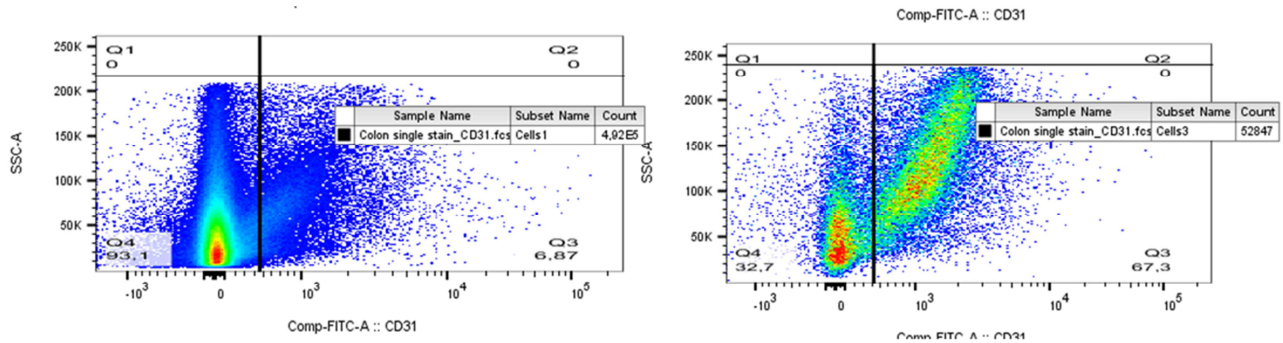


CD31

Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Very little fluorescence derived from CD31 was detected outside the unstained cell area.

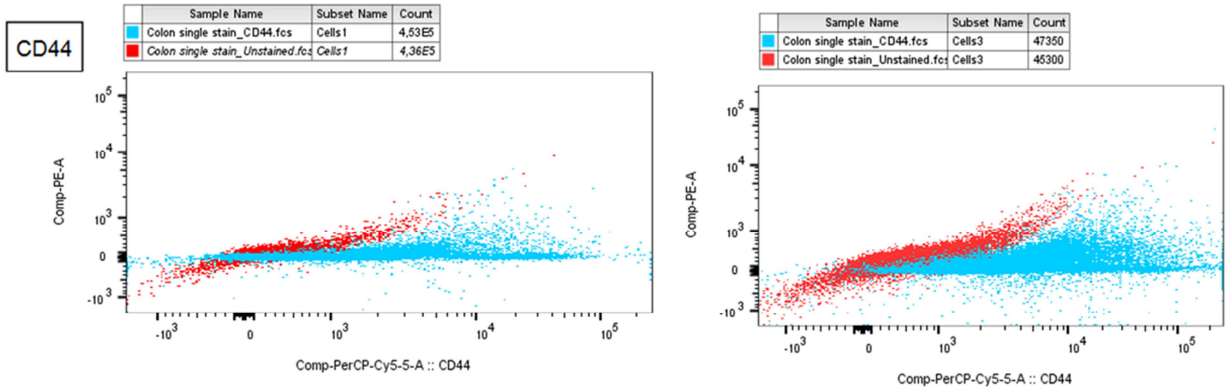


The scatter image of individual populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a CD31 positive population. The percent amount of events detected was 6,87 and 67,3 for groups 1 and 3 respectively.

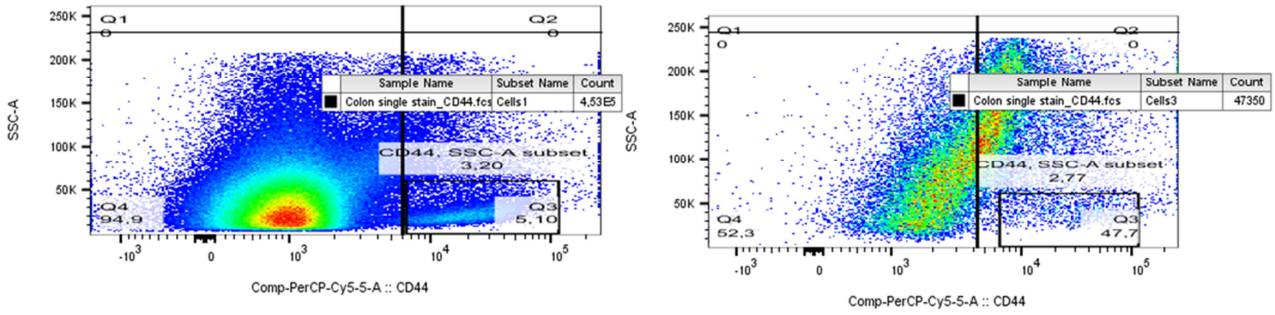


CD44

Groups 1 and 3 demonstrated a lesser overlap with the unstained cell group with a clearer defined delimitation of stained and unstained groups. The stained population is greater displaced to the right on a metric scala and possibly can be divided into two groups

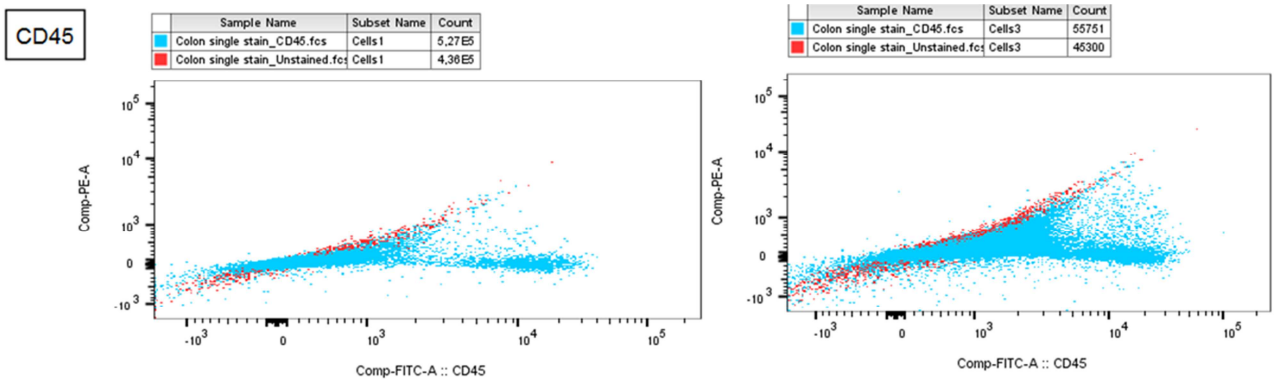


The scatter image of individual populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a CD44 positive population. The percent number of events detected was 5,10 and 47,7 for the separate population looking parts of groups 1 and 3 respectively.

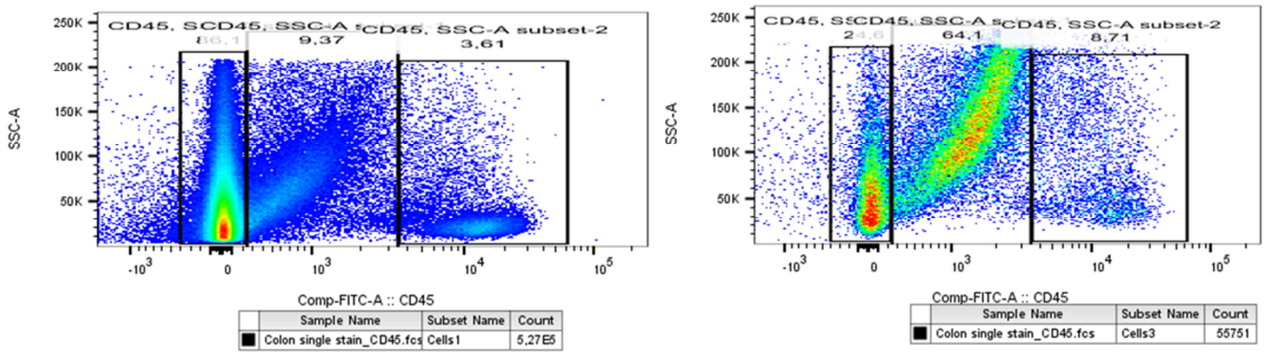


CD45

Groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. However the configuration of the stained area differs from the unstained area.

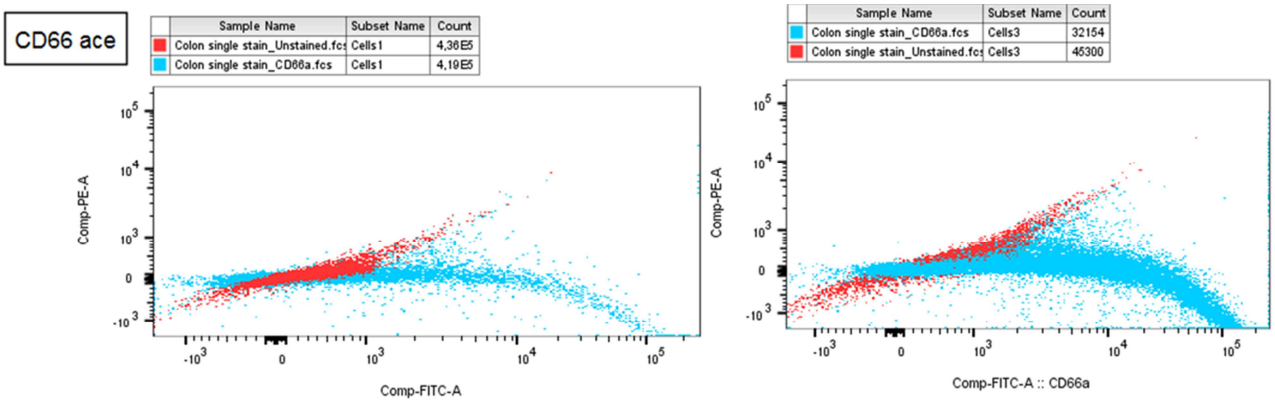


The scatter image of individual populations 1 and 3 demonstrates a population that can be divided into three groups. One of those, most likely the one to the far right can indicate CD45 positive target population. The percent number of events detected here was 3,61 and 8,71 for groups 1 and 3 respectively.

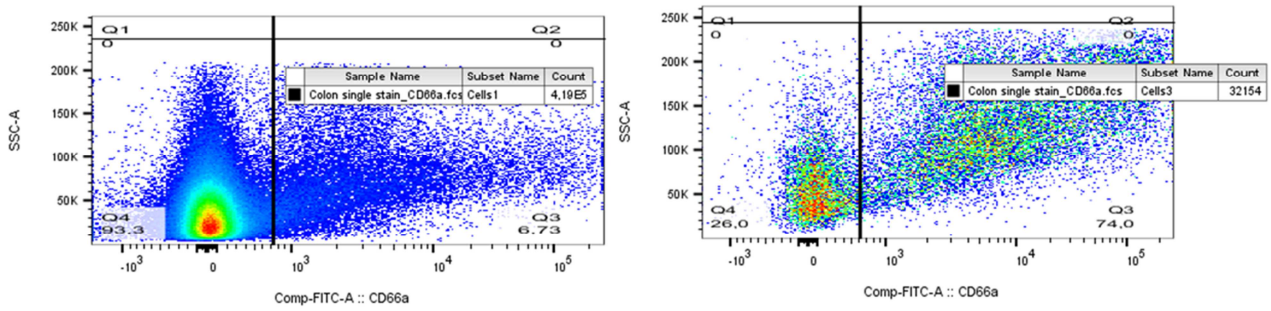


CD66c

Both groups 1 and 3 of the stained population demonstrated a certain overlap with the unstained cell group. The configuration of the stained area markedly differs from the unstained area.

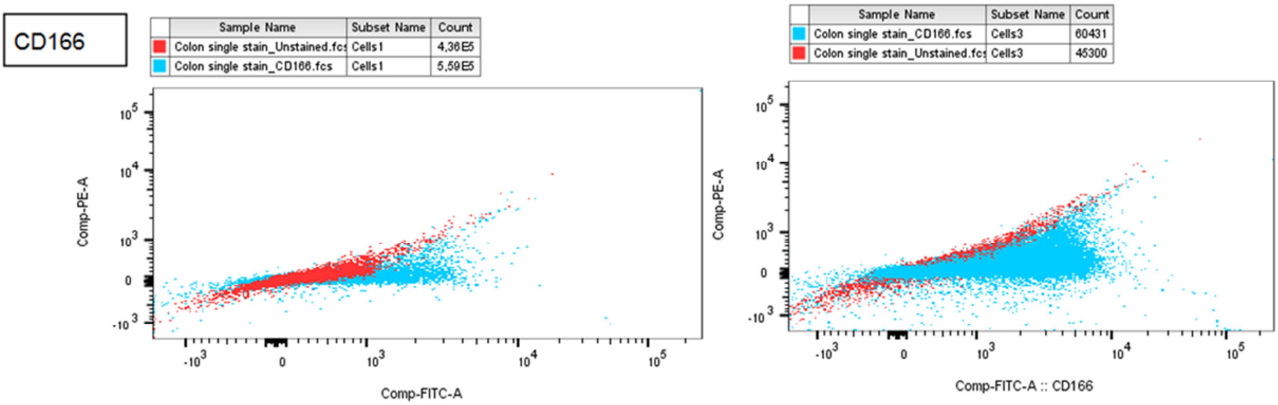


The scatter image of individual stained populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a CD31 positive population. The percent number of events detected here was 6,73 and 74,0 for groups 1 and 3 respectively.

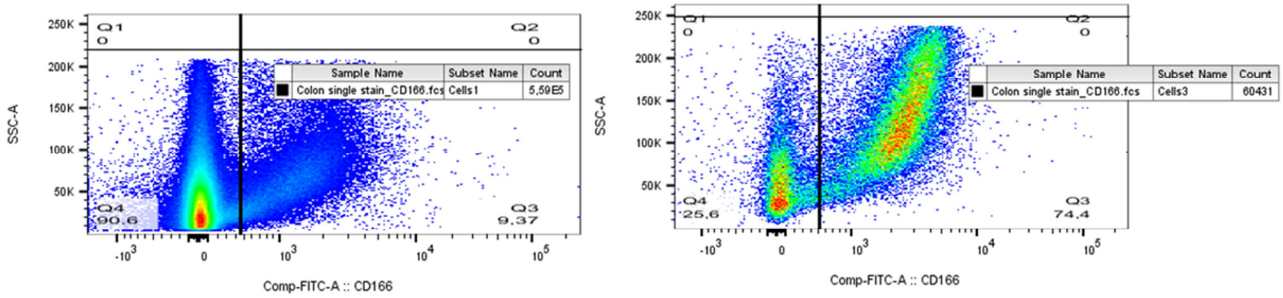


CD166

Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Only a little fluorescence derived from CD 166 was detected outside the unstained cell area.

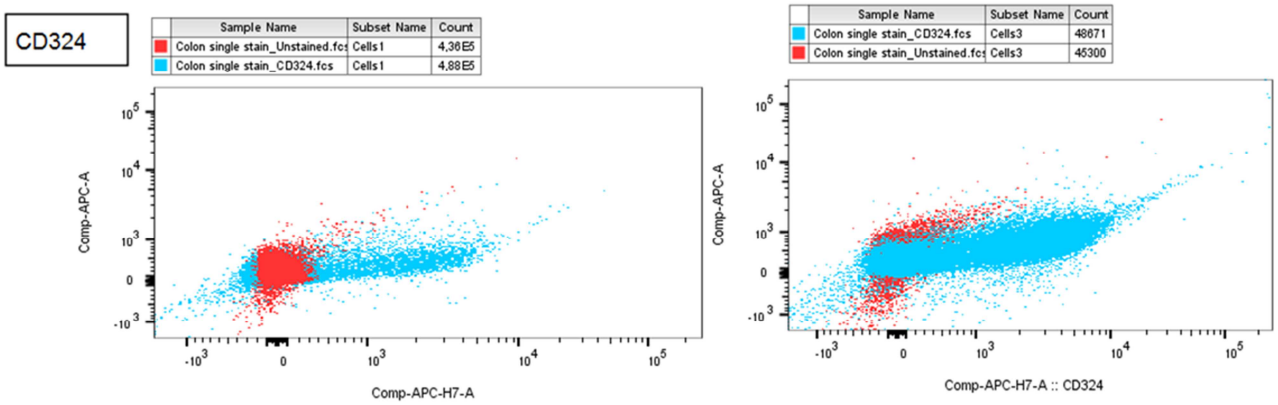


The scatter image of individual stained populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be CD166 positive population. The percent number of events detected here was 9,37 and 74,4 for groups 1 and 3 respectively.

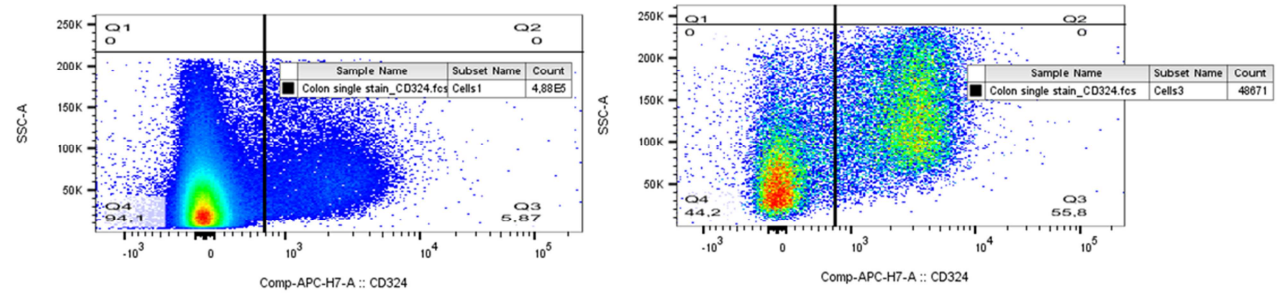


CD324

Both groups 1 and 3 demonstrated an overlap with the unstained cell group. Some fluorescence derived from CD324 was detected outside the unstained cell area.

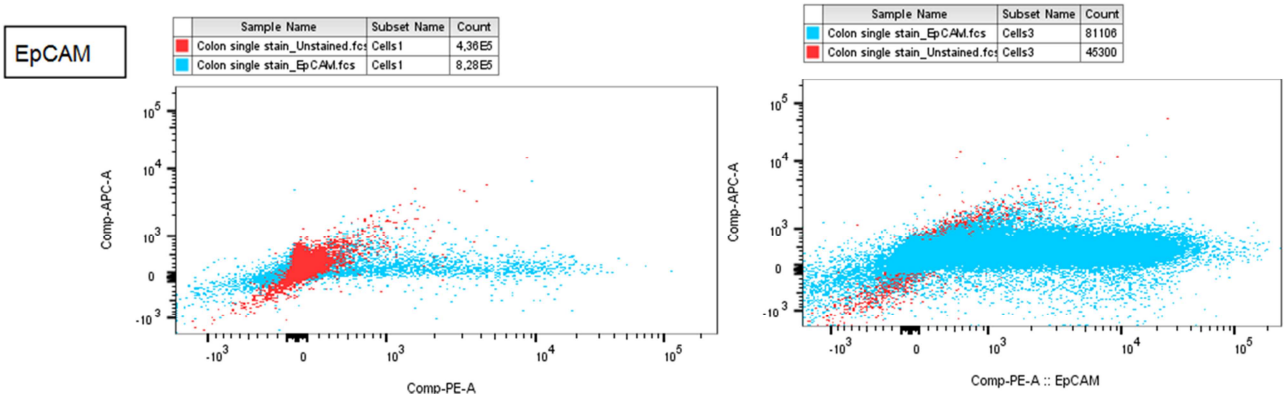


The scatter image of individual stained populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a CD323 positive population. The percent number of events detected here was 5,87 and 55,8 for groups 1 and 3 respectively.

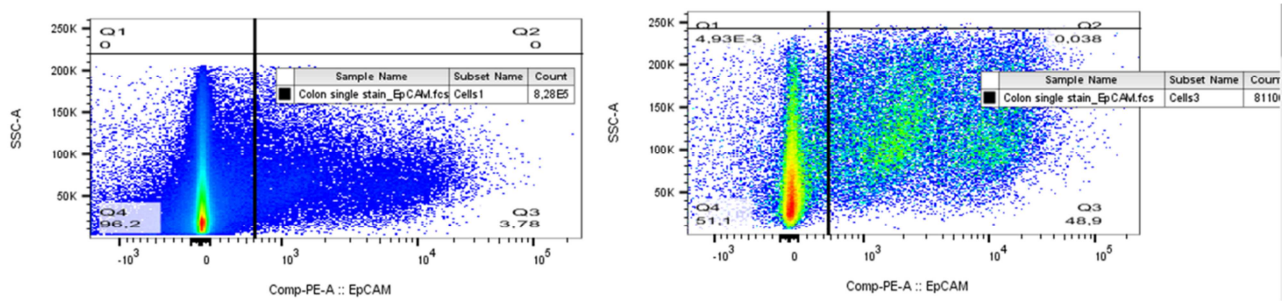


EpCAM

Both groups 1 and 3 demonstrated an overlap with the unstained cell group. However the configuration of the stained group differes substantially from the configuration of unstained group with a significant amount EpCAM derived fluorescens detected outside the unstained cell area.

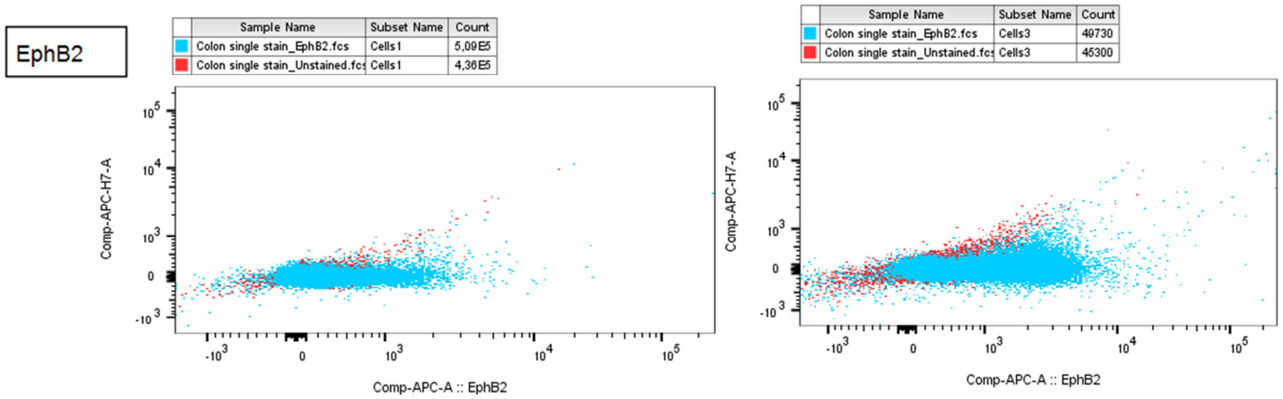


The scatter image of individual stained populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a EpCAM positive population. The percent number of events detected here was 3.78 and 48.9 for groups 1 and 3 respectively.

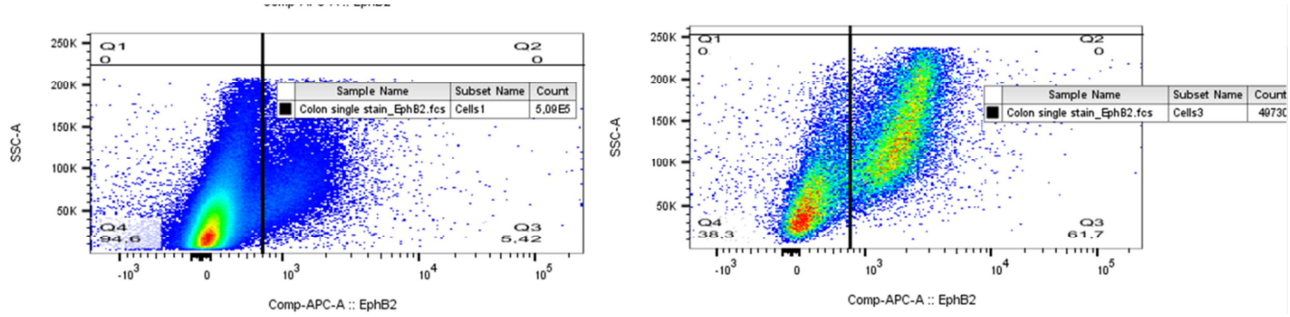


EphB2

Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. Very little fluorescence derived from EphB2 antibodies was detected outside the unstained cell area.

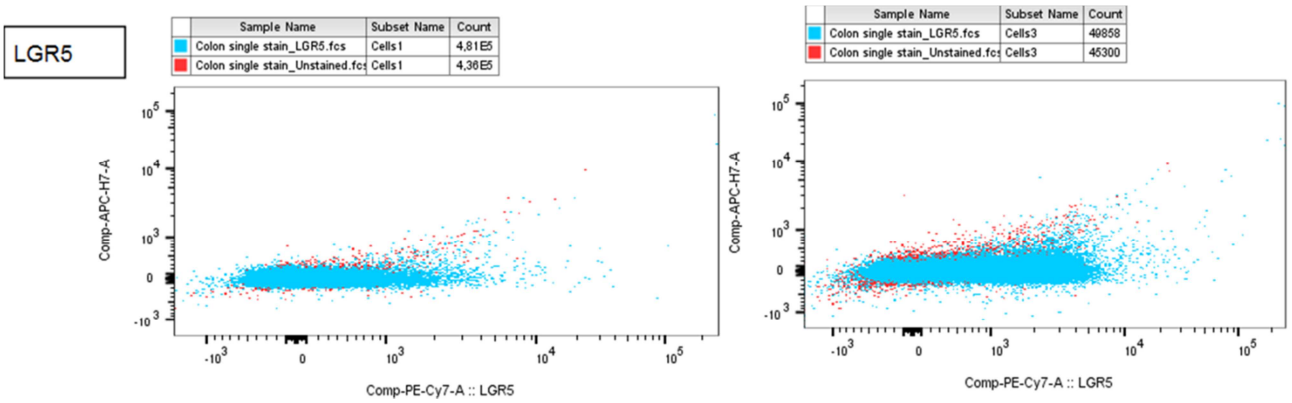


The scatter image of individual populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be a EphB2 positive population. The percent number of events detected here was 5,42 and 61,7 for groups 1 and 3 respectively.



Lgr5

Both groups 1 and 3 demonstrated a substantial overlap with the unstained cell group. A little fluorescence derived from Lgr5 antibody was detected outside the unstained cell area.



The scatter image of individual populations 1 and 3 demonstrates a population that can be divided into two groups. The one in the lower right quadrant might be Lgr5 positive population. The percent number of events detected here was 5,37 and 50,2 for groups 1 and 3 respectively.