REGULATION OF STEMNESS AND DIFFERENTIATION IN COLORECTAL CANCER

SHAAN-CHIRAG CHANDRAHAS GANDHI

Magdalen College

A thesis submitted to the Division of Medical Sciences, University of Oxford, in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Medical Oncology

Trinity Term 2010

Weatherall Institute of Molecular Medicine

University of Oxford

This thesis is dedicated to my parents, who are a source of never-ending support.

Abstract

The cancer stem cell (CSC) model of carcinogenesis and progression posits that within a tumor lies a subpopulation of cells that solely possess the ability to initiate a tumor and to differentiate into tumor cell lineages. Although the behavior of such cells is known, the challenge is to identify factors that characterize the CSC subpopulation. In this thesis, cell lines were identified that, when grown in three-dimensions, gave rise to organized colonies containing lumens originating from differentiating cells ("lumen lines") and to densely-packed, spherical colonies originating from non-differentiating cells ("dense lines"). A microarray comparison of the pair identified genes upregulated in dense lines, including *CD55* and *BMI1*, and in lumen lines, including *CDX1* (Chapter 3). CD55 was used to isolate $CD55^{high}$ CSCs via flow cytometry that are able to self-renew, differentiate, initiate more colonies, proliferate more rapidly and exhibit an increased G_2/M cell cycle population as opposed to unfractionated cells. Furthermore, the $CD55^{high}$ cells were able to give rise to more differentiated, lumen colonies *in vitro*, indicating that CD55 enriches for cells possessing a capacity to differentiate, and were able to enrich the $CD24^{high}CD44^{high}$ putative CSC population further (Chapter 4). CDNA induction of *BMI1* and *CDX1* expression led to increased clonogenicity/proliferation and decreased clonogenicity/proliferation, respectively, and incorporation of a *CDX1* reporter construct into the SW1222 cell line identified $CDX1^{+}$ cells as a low-expressing population of CD55 (Chapter 5). Finally, co-culture of cell lines in an *in vivo*-like environment with intestinal myofibroblasts promoted the CSC population by enhancing clonogenicity, proliferation and expression of CD55 (Chapter 6). The results of this thesis implicate CD55 as a potent marker of colorectal cancer stemness, link the expression of BMI1 and CDX1 to cancer stemness and differentiation, respectively, and identify a role for the *in vivo* stem cell niche in maintaining the CSC population. [299 words]

Acknowledgements

I am grateful for the help and support of my supervisor, Professor Sir Walter Bodmer, as I navigated the challenging waters that are graduate studies in cancer biology. His guidance and advice in probing the literature, designing experiments and interpreting results have been invaluable.

Furthermore, I would like to thank the members of the Cancer and Immunogenetics Laboratory, without whom my time in the laboratory would have been filled with far more troubleshooting and far less experimenting. Special thanks go Shazad Ashraf, Sylvia Bartlett, David Bicknell, Karin Bracht, Carol Chan, Olymbia Dikomitou, Rachael Hancox, Ying Liu, Angie Nicholls, Fillitsia Ntouroupi, Jenny Wilding and Trevor Yeung. Additional thanks go to the members of the Weatherall Institute of Molecular Medicine core facilities, including Ann Atzberger, Kevin Clark and Craig Waugh from the Flow Cytometry Core Facility, and to Professors Tariq Enver and Ian Hickson, whose advice during the DPhil transfer process was of great use.

Finally, I would like to thank my financial supporters, Cancer Research UK and the Rhodes Trust, without whose funds my work would not be possible.

Declaration

I, Shaan-Chirag Chandrahas Gandhi, hereby declare that the work on which this thesis is based is my original work and that neither the whole work nor any part of it has been, is being or is to be submitted for another degree at this or any other university.

SIGNED: The signature has been removed from this electronic version of the thesis DATE: 22 January 2011

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 The colon and the rectum

1.1.1 Anatomy

The colon and rectum form the final section of the human gastrointestinal tract. Measuring up to 1.5 meters, the large intestine, of which the cecum, colon and rectum are segments, functions mainly as a site for the storage of digested food wastes, the retention of water and the culture of gut bacteria to assist in further digestion and vitamin absorption (Aigner and Fritsch, 2010). Within the colon and rectum are eight individual sections corresponding to different attachment and connection geometries. In order, they are the ascending colon, right colic flexure, transverse colon, left colic flexure, descending colon, sigmoid colon, rectum and anal canal. The first six correspond to the colon and the last two correspond to the rectum (Aigner and Fritsch, 2010).

The short ascending colon connects to the transverse colon at the right colic flexure. The longer transverse colon follows the course of the pancreas, small intestinal duodenum and stomach, and connects not only to the descending colon via the left colic flexure but also to the greater omentum peritoneal tissue fold. The descending colon passes superior to the sigmoid colon, whilst the sigmoid colon curves slightly superior to connect to the rectum (Aigner and Fritsch, 2010). Macroscopically, the length of the colon is characterized by the presence of regular haustra, or colonic pouches, and smaller tissue folds within the haustra, while the rectum possesses few of these anatomic features (Aigner and Fritsch, 2010).

1.1.2 Histology

1.1.2.1 Crypt architecture

The basic unit of the colorectal epithelium is the crypt (Figure 1.1), an invagination of the continuous gut lining that mediates the main secretory and absorptive functions of the large intestine (Geibel, 2005). The millions of crypts present in a normal adult large intestine markedly increase the luminal surface area over a similar, but smooth, lining, enhancing the functional abilities of the large intestine.

At the base of the crypt lie the colonic stem cells, which differentiate upwards to regenerate the colonic lining. These stem cells initially differentiate into transit-amplifying cells, which lie directly superior to the stem cells and are thought to possess a greater proliferative ability (Humphries and Wright, 2008). These transit-amplifying cells divide and differentiate upwards even further into the three major differentiated epithelial lineages of the crypt, viz. columnar, goblet and enteroendocrine cells. The columnar cells are mainly responsible for absorptive functions, while the goblet and enteroendocrine cells secrete mucus and hormones, respectively (Humphries and Wright, 2008). Columnar and goblet cells make up 95% of the total crypt cell population, while enteroendocrine cells making up the remaining 5% (Geibel, 2005). Surrounding the crypt and bound to the basal lamina is a sheath of myofibroblasts, cells of a mesenchymal lineage possessing elements of both fibroblasts and smooth muscle cells. These cells are thought to be important in maintaining the colonic stem cell niche and in regulating the growth of the crypt through the secretion of growth factors (Humphries and Wright, 2008).

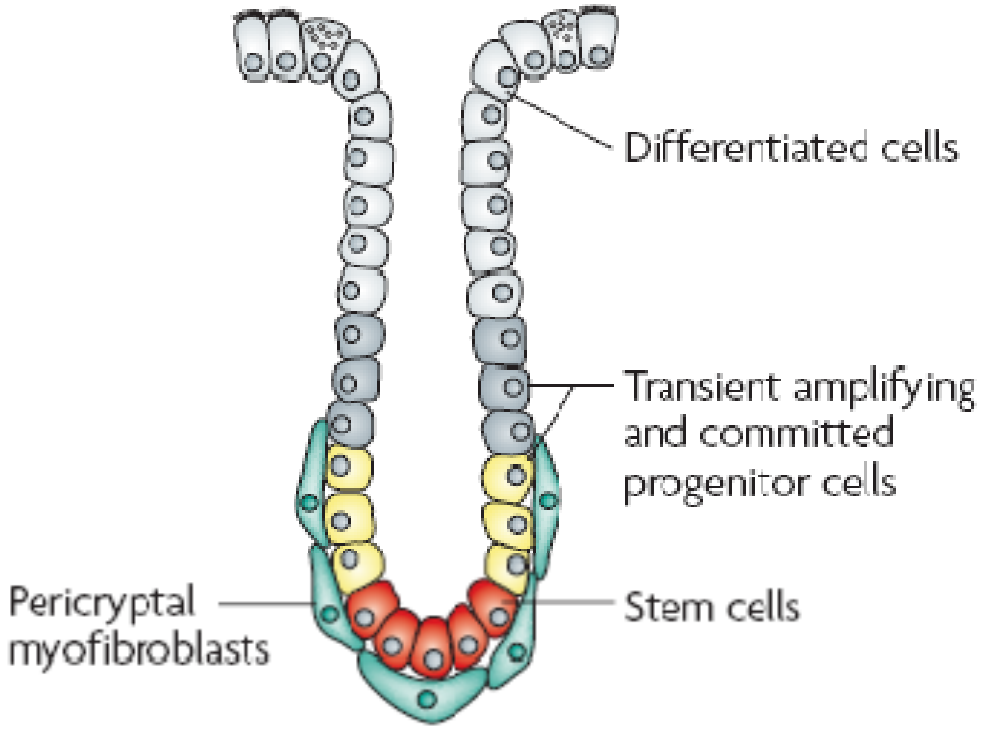


Figure 1.1 Structural organization of the colonic crypt, with stem cell at the base, transit-amplifying cells in the midsection, differentiated cells at the top and myofibroblasts lining the entire crypt length beneath the basement membrane (Humphries and Wright, 2008).

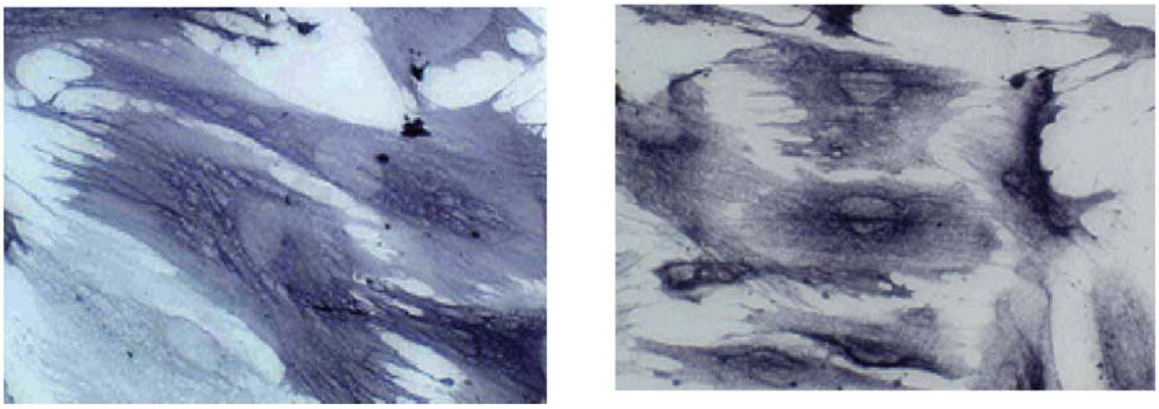


Figure 1.2 Immunostaining of myofibroblasts shows their ultrastructural ovoid/stellate hybrid features as well as their positivity for (left) α-smooth muscle actin and (right) vimentin (Andoh, et al., 2005).

1.1.2.2 Intestinal stem cells

The high rate of gut lining regeneration (Kim and Shibata, 2002) in the colon lends itself to the conclusion that stem cells must reside within crypts to drive cell division forward. Thus, multiple studies conducted over the past two decades have focused on proving the clonal nature of colonic crypts as well as on identifying the stem cells themselves. Novelli and colleagues (1996) found an XY/XO chimeric patient in which cells making up individual crypts either all possessed both X and Y chromosomes or only an X—no crypts were found to contain both karyotypes, indicating that individual crypts must be monoclonal. Further work identified a group of Sardinian women with intermittent expression of the glucose-6-phosphate dehydrogenase gene (*G6PD*) via X-inactivation—analysis of colonic tissue found numerous crypts with either normal G6PD activity or ablated activity, with none possessing cells from both activity populations (Novelli, et al., 2003). Immunohistochemistry of crypts possessing either wild-type copies of cytochrome *c* oxidase (COX) or mutant copies uncovered crypts either possessing only wild-type genes, only COX-mutant genes or a combination of both (Greaves, et al., 2006). The emergence of COX-mutant-only crypts indicates that crypt growth and division is clonal, and the presence of COX-mutant patches within the colon indicates crypt fission as the main avenue of colonic growth (Greaves, et al., 2006). Thus, it is clear that colonic crypts arise from the clonal expansion and differentiation of stem cells.

The specific identity and position of these stem cells has been a more complex challenge. Immunohistochemistry experiments analyzing expression patterns of caudal type homeobox transcription factor 1 (CDX1), an intestinal transcription factor known to be crucial in intestinal differentiation, found that the bases of human crypts expressed little or no CDX1, while the luminal edges of crypts expressed higher levels (Chan, et al.,

2009), implying a directional process of crypt differentiation emanating from potential $CDX1^-$ cells at the bases of crypts. Studies have also focused on positively identifying colonic stem cells. Barker and colleagues (2007) have identified populations of leucine-rich repeat-containing G protein-coupled receptor 5 (LGR5) (also known as GPR49)-expressing cells in the murine small and large bowel that possess the ability to generate full crypts. Transgenic mouse models expressing LacZ-tagged LGR5 under control of a tamoxifen promoter show the gradual dispersal of blue clones emanating from the bases of crypts after promoter induction—60 days after the initial tamoxifen pulse, uniformly blue crypts were visualized (Barker, et al., 2007). Furthermore, Sato and colleagues have demonstrated the ability of $LGR5^+$ cells to give rise to complex, organotypic structures strongly reminiscent of small intestinal crypts when cultured in a three-dimensional matrix (2009). In human colonic tissues, immunostaining of human crypt tissue samples has highlighted the RNA-binding protein Musashi-1 as a potential marker of colonic stem cells—Nishimura and colleagues (2003) found Musashi-1 expression to be primarily localized in the bottom segments of crypts. Thus, not only has the location of colonic stem cells been identified, but also their specific phenotype has begun to be ascertained.

1.1.2.3 Myofibroblasts

As discussed previously, surrounding the epithelial colonic crypt and directly inferior to the epithelial basement membrane is a sheath of pericryptal myofibroblasts that merges with the lower pericytes surrounding blood vessels to form a continuous subepithelial support layer (Powell, et al., 1999a). Derived from a mesenchymal lineage, myofibroblasts (Figure 1.2) possess features common with fibroblasts and smooth muscle cells, in that their morphology is an ovoid-stellate hybrid and that they are positive for both vimentin and α-smooth muscle actin protein expression (Andoh, et al., 2005). Since

they are positioned proximal to the crypt, it is likely that myofibroblasts are crucial in the maintenance of normal crypt growth and differentiation as well as the intestinal stem cell niche (Yen and Wright, 2006). Wnt signaling is important in regulating epithelial cell proliferation, and subepithelial myofibroblasts have been shown to secrete an array of Wnt proteins *in vitro*, implying a role in supplying proper Wnt ligand signaling to intestinal stem cells (Andoh, et al., 2005). Indeed, stimulated secretion of Wnt5a in the colonic myofibroblast cell line CCD-18Co can trigger the expression of the colon differentiation protein caudal type homeobox transcription factor 2 (CDX2) (Pacheco and Macleod, 2008), indicating a role for growth control on the part of myofibroblasts. Consequently, it is likely that such myofibroblasts can contribute to the aberrant cell growth and division events characteristic of cancer. Myofibroblast-conditioned medium promote migration and invasion of the T84 colorectal cancer cell line (McKaig, et al., 1999), while such medium can promote an undifferentiated state in primary colorectal cancer stem cells (Vermeulen, et al., 2010). Breast-associated fibroblasts can induce increased clone formation within the UACC-812 breast cancer cell line (Samoszuk, et al., 2005). Experiments with cancer-associated fibroblasts (CAFs), or fibroblasts that are specifically proximal to solid tumors (of which myofibroblasts can be a subset), have revealed their abilities to secrete pro-tumorigenic growth factors, including epithelial growth factor (EGF) and transforming growth factor-β (TGF-β) (Ostman and Augsten, 2009). Furthermore, CAFs can stimulate metastasis, chemoresistance and angiogenesis (Ostman and Augsten, 2009). Clearly, a strong role exists for associated fibroblasts in regulating tumor promotion—it is likely, therefore, that crypt-associated myofibroblasts play a role in colorectal cancer.

1.2 Colorectal cancer

1.2.1 Carcinogenesis

1.2.1.1 Morphological changes

The process of colorectal carcinogenesis (Figure 1.3), or the formation of a neoplastic lesion within the epithelial layer of the colon or rectum, begins with a larger-than-normal rate of cell division mediated by the crypt intestinal stem cells and transit-amplifying cells. The subsequent expansion of crypt epithelial cells leads to the formation of aberrant crypt foci (ACF), potential proto-tumors consisting of either benign or malignant cells. Certain ACFs can grow further to yield polyps, or unorganized masses of cells that protrude into the colon luminal cavity. Dysplastic polyps, also known as adenomas, can proceed further into carcinomas. These highly abnormal and proliferative cellular architectures have lost attachment to and control by the basement membrane and have invaded into surrounding tissues (Fodde, et al., 2001). The process from ACF to adenoma to carcinoma is not an inexorable one—multiple decision points occur on the tree where genetic and epigenetic changes are necessary to drive the further growth and disorganization of the colonic neoplasm.

1.2.1.2 Genetic changes

The transitions from normal colonic epithelium to an aggressive, invasive carcinoma require multiple losses of the normal genetic controls on cellular growth and division. In the case of colorectal cancer, evidence points to the gene *APC* (adenomatous polyposis coli) as the key initiator of ACFs. A Wnt target tumor suppressor gene that generates a 312 kilodalton (kDa) protein, *APC* is the most ubiquitously-mutated gene in colorectal

cancer and is thought to be the first gene to be mutated in the multi-step progression to a carcinoma (Powell, et al., 1992). APC acts by sequestering β-catenin, a Wnt pathway transcription factor, from the nucleus in the absence of a Wnt ligand—nonfunctional APC is unable to bind β-catenin in the cytoplasm, allowing the transcription factor to localize to the nucleus and promote transcription of pro-growth genes in the absence of a ligand stimulus (Korinek, et al., 1997). Most *APC* mutations are clustered between codons 1286 and 1513 and produce truncation protein products with few or no β-catenin binding/regulatory regions or binding motifs to axin or conductin, two other proteins critical for the tumor suppressor functioning of APC (Miyaki, et al., 1994). *APC* was initially identified as the first mutated gene in the pathway toward colorectal carcinogenesis in the most common hereditary form of bowel cancer, familial adenomatous polyposis (FAP), a form responsible for up to 5% of colorectal cancer cases (Kinzler and Vogelstein, 1996). Furthermore, the most commonly used mouse model of colorectal cancer, the Apc^{Min} model, is characterized by its expression of an APC truncation protein product of size 95 kDa (Fodde and Smits, 2001).

After loss of proliferative control by APC, activating mutations in the oncogene *KRAS*, generally in codon 12 of the gene, likely drive further cell division and growth to transform ACFs into more serious adenomas (Bos, et al., 1987), although only approximately 50% of colorectal tumors possess *KRAS* mutations, suggesting a role for other oncogenes or non-oncogenic pro-growth pathways (Fodde, et al., 2001) in the initiation of adenomas. Further losses of tumor suppressor genes, including the TGF-β pathway component *SMAD4* as well as *TP53*, prime adenomas to develop into carcinomas (Fodde, et al., 2001). Concurrent with the two-hit mutation model of Knudson (1971), requiring two mutations in both alleles of tumor suppressor genes (i.e. loss of heterozygosity, or LOH), it is likely that at least seven specific mutations (two loss-of-

function mutations each in *APC*, *SMAD4* and *TP53* and one gain-of-function mutation in *KRAS*) (Fodde, et al., 2001) are required for the growth of a colorectal tumor, underscoring the importance of genetic changes in colorectal carcinogenesis. Indeed, LOH analysis has uncovered three regions with particularly high incidences of mutation, namely 5q, 17p and 18q (Vogelstein, et al., 1988), which correspond to the chromosomal loci of *APC*, *TP53* and *SMAD4*, respectively.

At the same time, LOH may not necessarily be linked to gene alterations found in colorectal cancer—Renkonen and colleagues (2005) found a group of families suffering from FAP that not only tested negative for *APC* mutations, but upon sequence analysis were found to harbor losses of heterozygosity in the *APC* 5q chromosomal region. LOH analysis of the fourth-most mutated region, 8q, uncovered 21 potential lost genes, although sequencing confirmed the presence of only fourteen genes within the lost alleles. Interestingly, assessment of tumor biopsies found no change in protein expression of genes found within the 8q LOH region, implying a more complex regulatory pattern for colorectal cancer-associated mutations (Mourra, et al., 2008).

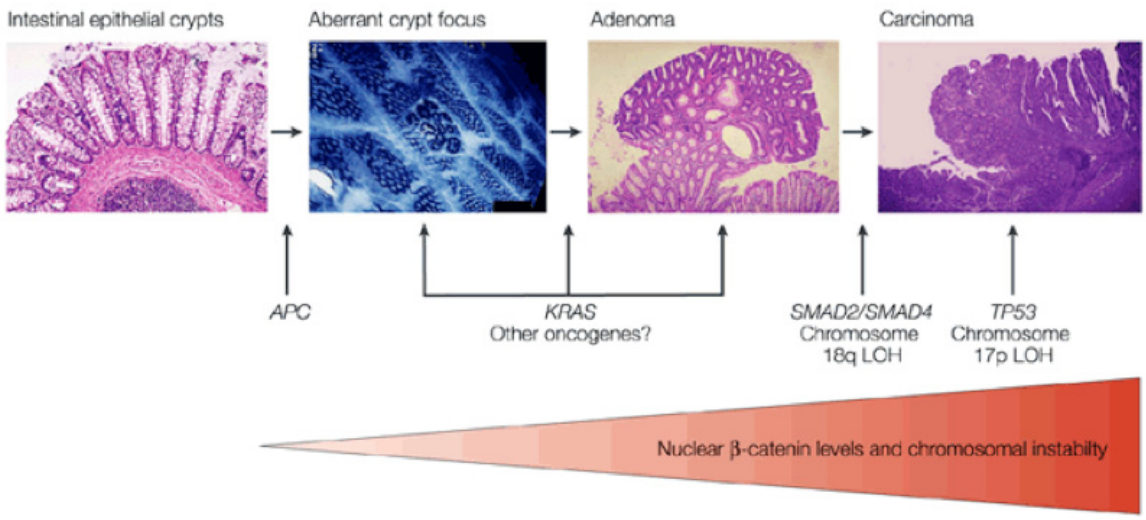


Figure 1.3 The morphological progression of colorectal cancer is highly regulated by mutations in certain key genes—*APC* is important in the transition between normal colonic epithelia and aberrant crypt foci (ACF), while *KRAS* and other oncogenes are constitutively activated after the loss of *APC* tumor suppression to drive ACFs forward toward adenomas; further losses of tumor suppressor gene activity, including *SMAD2*, *SMAD4* and *TP53* stimulate the formation of more malignant carcinomas (Fodde, et al., 2001).

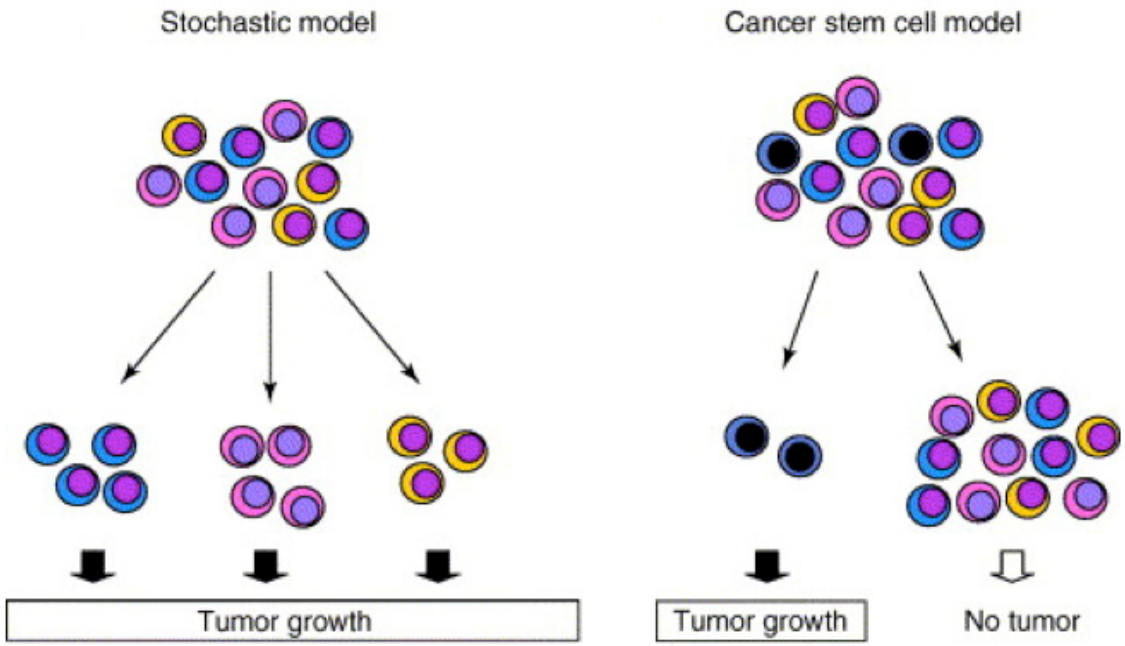


Figure 1.4 The two dominant models of tumor initiation are the (left) stochastic model, characterized by the ability of all cells within a tumor to give rise to a new cancer, and the (right) cancer stem cell model, characterized by the presence of a subpopulation which solely possesses the ability to initiate tumors (Wang and Dick, 2005).

1.2.2 Incidence

Colorectal cancer represents a significant source of cancer-related morbidity and mortality throughout the world. In the United Kingdom alone in 2006, 37,514 new cases of colon and rectal cancers were diagnosed (CRUK, 2010a) (an incidence of 45.3 cases per 100,000 individuals) (CRUK, 2010b), making the pair the fourth-most diagnosed malignancy after non-melanoma skin, lung and breast cancers. Furthermore, in 2007, colorectal cancer claimed the lives of 16,007 Britons (CRUK, 2010a) (an incidence of 17.7 deaths per 100,000) (CRUK, 2010b), making it the second-leading cause of cancer death, again after lung cancer. These statistics are not unique to the UK—colorectal cancer is estimated to be responsible for 146,970 new cases and 49,920 deaths in 2009 in the United States, making it the fourth-most diagnosed cancer and the second-leading cause of cancer death in that country (Jemal, et al., 2010). Colorectal cancer tends to afflict men more than women (CRUK, 2010b) and tends to increase in incidence with age (Armitage and Doll, 1954). The incidence-to-age relationship could potentially relate to the number of genetic mutations thought to be necessary for colorectal carcinogenesis—with time comes the greater probability of genetic mutations to occur.

1.2.3 Treatment

Treatment for colorectal cancer is usually surgical resection and chemotherapy, although for metastatic cancer, chemotherapy is crucial in targeting diffuse and difficult-to-reach metastases. 5-Fluoruracil (5-FU), a thymidylate synthase inhibitor which targets actively-dividing cells, has been standard treatment for over forty years, although in recent years, it has been combined with other cytotoxic drugs, such as folinic acid and oxaliplatin (Slevin and Payne, 2004). However, as with all cell division inhibitors, the differential impacts on cancer versus normal cells are limited. The effects of such drugs are systemic,

and side effects are severe. New targeted technologies promise to eliminate cancer cells more efficiently and effectively by focusing on the biology of tumorigenesis, be it through small molecule inhibition of overexpressed tyrosine kinases (such as gefitinib or lapatinib) or monoclonal antibody binding (such as bevacizumab and cetuximab) (Slevin and Payne, 2004). Unfortunately, many tumors exhibit signs of chemotherapy resistance, reducing the effectiveness of such therapies (Boman and Huang, 2008).

1.3 Cancer stem cells

1.3.1 Definition

Since the process of colorectal carcinogenesis requires multiple genetic mutations and occurs over decades, as suggested by Armitage and Doll (1954), it is likely that the target of these mutations is a compartment of long-lived colonic cells, potentially the colonic stem cells. Accumulating evidence over the past five decades has lent credence to the hypothesis that, thus, not all cells within a tumor possess the ability to initiate or propagate the cancer. Rather, a subpopulation of biological distinct cells, deemed "cancer stem cells," or CSCs possesses carcinogenic properties, rather than all cells, as the prevailing stochastic model of carcinogenesis has predicted (Wang and Dick, 2005) (Figure 1.4). Formally, the American Association for Cancer Research (AACR) consensus definition of CSCs states that they are cells "within a tumor that possess the capacity to self-renew and to cause the heterogeneous lineages of cancer cells that comprise the tumor" (Clarke, et al., 2006).

1.3.2 Initial evidence

Much of the original work relating to the CSC model of tumorigenesis was focused on the hematopoietic system and its associated malignancies. The pioneering studies of Till and McCulloch (1961) on murine hematopoiesis found that irradiated mice transplanted with donated bone marrow possessed spleens containing clonal clumps of cells with diverse lineages. When these clumps were disaggregated and re-transplanted, only a subset of transplanted cells was able to recapitulate the clonal cell colonies containing red blood cells, white blood cells and platelets. These results demonstrated the presence of hematopoietic stem cells and their abilities to divide asymmetrically and to differentiate to give rise to the diversity of cell lineages in the normal bloodstream. Similar work with murine lymphoma cells came to a similar conclusion—there exists a subset of cells within the cancerous population with the ability to grow colonies possessing the various blood cell types (Bruce and van der Gaag, 1963).

Perhaps, though, the most persuasive early evidence supporting the presence of CSCs and the veracity of the CSC model comes from Hamburger and Salmon (1977). In their study, primary cells from lung, neuroblastoma and ovarian solid tumors were cultured in soft agar to assess anchorage-independent clonogenicity. Of the generated clones, only 1 in 1,000 to 1 in 5,000 cells within each clone were able to give rise to a colony in soft agar, suggesting that a smaller population of cancer cells within each tumor possessed the ability to initiate another colony *in vitro* (a surrogate for initiating a tumor *in vivo*).

A challenge faced by the previous studies was that populations of CSCs were identified retrospectively. More recent studies have begun to tackle the challenge of CSC prospective identification. Work by Bonnet and Dick (1997) found a subpopulation (0.2% of the total) of primary human acute myeloid leukemia cells that was $CD34^{++}CD38^{-}$ that

possessed the ability to give rise to new leukemias when injected into immunocompromised non-obese diabetic/severely compromised immunodeficient (NOD/SCID) mice. The xenograft experiments demonstrated the ability of these CSCs to initiate tumors *in vivo*. Furthermore, these $CD34^{++}CD38^{-}$ leukemia stem cells were able to give rise to leukemic blast cells, demonstrating their ability to differentiate or "give rise to the heterogeneous lineages of cancer cells" (from the AACR CSC definition).

The ability of scientists to use the cell surface phenotype of cells to identify CSCs before tumorigenic assessment has recently been studied in solid tumors as well. The identification (Al-Hajj, et al., 2003) of a subpopulation of primary breast cancer cells with the cell surface protein phenotype $CD44^{+}CD24^{-/low}Lineage^{-}$ (the lineage phenotype is a combination of the immunological proteins CD2, CD3, CD10, CD16, CD18, CD31, CD64 and CD140b) that not only initiated tumors in immunocompromised mice but also retained its phenotype upon serial *in vitro* passage demonstrated that the CSC model of carcinogenesis may be a valid model to describe tumor initiation and organization. These findings were extended to colorectal cancer in 2007, where concurrent work by O'Brien and colleagues as well as Ricci-Vitiani and colleagues found a subpopulation of $CD133^{+}$ cells in primary colorectal tumors capable of serial passage and initiating tumors in NOD/SCID mice.

1.3.3 Implications

1.3.3.1 Origins of cancer stem cells

Perhaps the most basic question regarding the CSC model focuses on their actual sources. Are CSCs originally derived from multipotent tissue stem cells that have lost the normal growth controls, or are they derived from abnormally-proliferating progenitor or

differentiated cells that have dedifferentiated? Since the hematopoietic system is the most well-characterized with respect to the identities and hierarchies of stem, progenitor and differentiated cells, most investigations of the cells of origin of CSCs have focused on leukemias and lymphomas. Although both hematopoietic stem cells and leukemic stem cells were found to share the same basic phenotype, $CD34^{+}CD38^{-}$, more detailed studies of the specific surface phenotype of leukemic stem cells have revealed their greater similarity with progenitor cells rather than with stem cells (Lobo, et al., 2007). However, other studies involving specific cellular manipulations implicate the stem cell population as the CSC originator—transduction of a mutant fusion gene product important in acute myeloid leukemia (AML) initiation into both stem cells and myeloid progenitor cells revealed that fewer stem cells were required to establish colonies than progenitor cells (Krivtsov, et al., 2006). Jamieson and colleagues (2004) found that an in-depth analysis of primary cell surface phenotypes of CSCs from chronic-phase chronic myeloid leukemia (CML) found them to be more in common with hematopoietic stem cells, while CSCs from blast crisis CML were found to be more similar to progenitor cells. If anything, the CSC population appears to be quite plastic, and the cell of origin may differ from tumor type to tumor type.

In solid tumors, because far less is known about the identity of the stem cell compartment, more work is remaining to ascertain the cell of origin of CSCs. In breast cancer, work by Shipitsin and colleagues (2007) found through microarray gene expression comparisons that the normal primary breast tissue stem cell population and primary breast CSC population were similar in surface phenotype ($CD44^{+}CD24^{-}$ $PROCR^{+}CD10^{-}$) and in mRNA expression. A more targeted study by Barker and colleagues (2009) with $LGR5^{+}$ putative primary tissue small intestinal stem cells discovered that the transformation of murine LGR5+ cells with a mutant *APC* gene copy

induced the formation of tumors, elegantly demonstrating the potential role for dysregulated small intestinal stem cells in giving rise to CSCs and/or to a tumor. Within human large intestinal cancers, the true cell of origin has still yet to be discovered.

1.3.3.2 Population number

The initial studies by Hamburger and Salmon (1977), Bonnet and Dick (1997) and Al-Hajj and colleagues (2003) all found the CSC population to be quite rare—the first two reports of colorectal CSCs (O'Brien, et al, 2007; Ricci-Vitiani, et al., 2007) found that $CD133^+$ putative CSCs make up less than 2.5% of the total population. Thus, a common inference drawn from these early studies was that the CSC population must be rare. Indeed, in many experiments using a variety of cancer cells, both from hematological and solid malignancies, the identified CSC population was indeed rare.

Less than 0.001% of primary melanoma cells were found to be $ABCB5^+$ CSCs (Schatton, et al. 2008), while in pancreatic cancer, the $CD133^+$ CSC population was found to less than 4% of the bulk primary cell population (Hermann, et al., 2007). However, rarity appears not to be a critical element of CSC physiology—numerous studies have found CSC frequencies to exceed 10% in certain tumors. Singh and colleagues (2004), in the first publication to identify CSCs in a neurological cancer, found that 19% of cells from primary glioblastoma tumors were $CD133^+$ CSCs. Indeed, in the C6 glioma cell line, all cells were found to possess tumor-initiating properties (Zheng, et al., 2007), while in melanoma, newer experiments using an extremely immunocompromised mouse model (NOD/SCID IL2Rγ^{null} mice) indicate that primary melanoma CSCs may be as frequent as 27% of the bulk cell population (Quintana, et al., 2008). Although some posit that cell populations with a high frequency of tumor-initiating cells actually support the stochastic model of tumorigenesis (Kai, et al., 2009), mathematical modeling of colonic crypt

dynamics by Johnston (2008) has indicated that the stem cell population can vary widely in number as a result of small variations in cellular growth rates and/or the presence of growth stimuli. Clearly, the CSC population can vary drastically in number from tumor to tumor, from rare to common.

1.3.3.3 Tumor heterogeneity

Perhaps the reasoning behind the inclusion of the phrase "stem cells" in the term "cancer stem cells" is that the process of differentiation CSCs are thought to undergo in order to give rise to the multiple cell lineages found in a tumor is not dissimilar to the process of differentiation normal multipotent stem cells undergo to give rise to different cells in tissues. Tumors are widely known to be heterogeneous in cell type, so if the CSC model of tumor initiation were to be true, then tumors formed from purified CSCs should not only possess the diverse cell types found in the primary tumor but also recapitulate the distribution and organization of the original tumor. Experimental evidence has supported this claim (Shackleton, et al., 2009). *In vivo* work with primary cells from breast (Al-Hajj, et al., 2003), brain (Singh, et al., 2004) and colorectal (O'Brien, et al., 2007; Ricci-Vitiani, et al., 2007) cancers has found that CSC-initiated tumors not only give rise to heterogeneous tumors but also that such heterogeneity reflects that of the primary tumor source of the purified CSCs (e.g. poorly-differentiated primary tumors generate CSCs that initiate poorly-differentiated murine secondary tumors). Isolated CSCs can also differentiate *in vitro* to give rise to complex structures such as cysts and branched tubules indicative of a multilineage potential, as demonstrated with the rat mammary adenocarcinoma cell line LA7 (Zucchi, et al., 2007). Work with hematological malignancies not only supports the notion of CSC-derived heterogeneous cancers but also the idea that such heterogeneity is hierarchically organized (Shackleton, et al., 2009).

Thus, a homogeneous CSC population is likely to possess the ability to create a heterogeneous bulk tumor, behavior consistent with the "stem cell" appellation.

1.3.3.4 Treatment resistance

The CSC model of carcinogenesis identifies the CSC population as the sole initiator and driver of carcinogenesis and, implicitly, argues that the molecular biology of CSCs is distinct from that of non-CSCs. Thus, from a treatment perspective, it is crucial that therapies designed against cancer be attuned to targeting the true originators of the tumor. At the same time, if cancer therapies were unable to eliminate the CSC population, tumor reoccurrence would be the result. Evidence from clinical oncology has supported this notion. Clinicians often must vary the chemotherapy or radiotherapy treatment regime to avoid tumor resistance (Boman and Huang, 2008). Furthermore, even after a tumor has gone into remission, observed recurrences force patients to undergo exhaustive therapy, often with toxic side effects.

The challenge, therefore, is to understand the impact of extant drug and radiation therapies on CSCs and to develop new CSC-targeted modalities to reduce treatment resistance, tumor reoccurrence and treatment adverse effects. Multiple studies on CSCs have found such cells to be more resistant to standard cytotoxic drugs. Nakai and colleagues (2009) found that glioblastoma cell line CSCs are more resistant to doxorubicin, etoposide and carboplastin than non-CSCs. Ovarian cell line CSCs are more resistant to cisplatin and paclitaxel than non-CSCs (Zhang, et al. 2008). Studies have also linked CSCs to radiation therapy resistance. In the breast cancer cell lines MCF7 and MDA-MB-231, $CD44^{+}CD24^{-/low}$ CSCs were resistant to radiation therapy versus non-CSCs (Phillips, et al., 2006). Not only are CSCs resistant to therapy, but also their populations are enriched following drug or radiation treatment. In the colorectal cancer

cell line HT29, treatment with 5-FU or oxaliplatin (two of the three pillars of standard colorectal cancer chemotherapy) enriched the proportion of $CD133^+$ CSCs more than sixteen-fold and the proportion of $CD44^+$ CSCs by double (Dallas, et al., 2009). Long-term treatment of an osteosarcoma cell line, MG-63, with the drug 3-aminobenzamide also selectively enriched the CSC population within those cells, yielding a new cell line (3AB-OS) with high expression of the potential CSC markers ABCG2 and CD133 (Di Fiore, et al., 2009). The same observations have recently been made using the SW620 colorectal cancer cell line: a 15-week treatment with cisplatin or doxorubicin triggered an expansion of the MDR^+ CSC population (Bartkowiak, et al., 2009).

Clearly, there is room in the oncologic armamentarium for better therapies that either target CSCs preferentially or at least decrease the abilities of CSCs to evade treatment. Todaro and colleagues (2007) found that $CD133^+$ primary colorectal CSCs are less able to resist 5-flurouracil and oxaliplatin cytotoxicity when treated with a monoclonal anti-interleukin-4 (IL-4) antibody, indicating a role for IL-4 in mediating CSC chemoresistance. The aforementioned study of chemoresistance in HT29 cells found that treatment with an insulin-like growth factor 1 receptor (IGF-1R) monoclonal antibody enhanced the cytotoxic effects of 5-FU and oxaliplatin (Dallas, et al., 2009). Even more interestingly, Hoey and colleagues (2009) recently reported positive results regarding a novel monoclonal antibody against delta-like protein 4 (DLL4), a component of the Notch pathway—the antibody, OMP-21M18, specifically targeted CSCs in breast and colorectal cancer. The antibody is now in a Phase I clinical trial to ascertain its clinical usefulness.

However, treatment resistance may not be a characteristic of all CSCs—Al-Assar and colleagues (2009) found that most isolated CSCs from a panel of breast and pancreatic

cancer cell lines were no more resistant to radiation than non-CSCs, while Aulmann and others (2010) found from a panel of primary breast cancer biopsies that the $CD44^+CD24^-$ CSC population decreases, rather than increases, after treatment with epirubicin and cyclophosphamide. Perhaps these results are not so much a repudiation of the idea that resistance phenotype is associated with CSCs but rather an indication that CSC purification techniques need to be refined.

1.3.3.5 Metastasis

The process of metastasis involves a cell or cells escaping a primary tumor site, traveling via the circulatory or lymphatic system and initiating a distal secondary tumor. This disseminating ability could potentially be the result of epithelial cancer cells undergoing an epithelial-mesenchymal transition (EMT), a process that involves the transformation of the epithelial cell phenotype into one that is mesenchymal, unpolarized and more able to escape the control mechanisms exerted by the epithelial basement membrane (Geiger and Peeper, 2009). Since the CSC model of carcinogenesis posits that CSCs are the cells solely able to initiate tumors, it is natural to hypothesize that metastasis-mediating cells are, in fact, escaped CSCs that may have undergone EMT. There is some evidence that has supported such a claim. Work in head and neck squamous cell carcinoma tumor biopsies (Chen, et al., 2009) with aldehyde dehydrogenase 1-positive ($ALDH1^+$) CSCs found that such CSCs highly expressed Snail, known to be a protein mediator of EMT, and that small interfering RNA (siRNA) knockdown of Snail expression ablated ALDH1 expression and blocked tumorigenicity of the isolated CSCs in mouse xenografts. Santisteban and colleagues (2009) were able to induce EMT of injected human breast cancer cell lines via an immunological response in *neu*-Tg mice. What was interesting about the induction of EMT was that the cells possessed a $CD44^+CD24^-$ phenotype,

identical to the phenotype thought to be associated with breast CSCs, and that the cells were able to initiate additional tumors, a direct link between EMT induction and CSCs. More circumstantially, the genes *EZH2* (enhancer of zeste homolog 2) and *BMI1* (B lymphoma Mo-MLV insertion region 1), key in stem cell maintenance pathways, are often overexpressed in metastatic cancers (Geiger and Peeper, 2009).

However, it could be possible that only a subset of the CSC compartment possesses the ability to metastasize. Hermann and colleagues (2007) found that although all $CD133^+$ CSCs isolated from primary pancreatic tumors were able to form tumors in immunocompromised mice, only the $CD133^+$ subset expressing CXCR4 (chemokine C-X-C motif receptor 4), a chemokine receptor important in hematopoietic stem cell homing, generated tumors that metastasized. It is even possible that the metastases possess few or no CSCs, as defined by cell surface phenotype—a study of human colorectal cancer in NOD/SCID mice revealed that although primary patient tumors express CD133, a putative marker of colorectal cancer stemness, metastatic secondary patient tumors could be initiated by both $CD133^+$ and $CD133^-$ cells (Shmelkov, et al., 2008). Although the parallels between CSCs and metastasis are tantalizing, there is clearly more work that needs to be completed to ascertain the specific relationship between the two.

1.3.4 Common experimental tools

The rigorous isolation and characterization of CSCs is a major goal of the cancer stem cell biology scientific community. The most commonly used method for prospective isolation of CSCs is cell sorting, either by fluorescence activation (FACS) or magnetic activation (MACS). In FACS, cancer cells are treated with a primary antibody raised against a transmembrane or surface protein; this antibody is frequently conjugated to a

fluorescent molecule, such as fluorescein isothiocyanate (FITC), phycoerythrin (PE) or allophycocyanin (APC). Alternatively, a secondary antibody conjugated to a fluorescent molecule can be used to label the protein of interest. Cells are analyzed by a flow cytometry apparatus and separated according to level of fluorescence. Positively- or negatively-expressing cells can be isolated by comparison to a control population (treated with a non-human-reacting isotype control antibody) or by sorting a fixed percentage of the fluorescence distribution. A similar procedure is used for MACS, but instead of a fluorescent secondary antibody, a secondary antibody conjugated to a magnetic bead is used. Treated cells are passed through a magnetic sieve—positively-expressing cells remain, while negatively-expressing cells pass through. Although MACS allows less control over the specific protein expression level of the isolated cells, its cell capacity is higher, allowing for high-throughput applications. Once putative CSCs are isolated by surface protein phenotype, a variety of *in vitro* experiments can be conducted to verify their activity. Hamburger and Salmon (1977) used growth of cells in soft agar (a gel-like substance with less than 1% dissolved agar) to test for anchorage-independent growth and ability to form colonies, a known *in vitro* surrogate for *in vivo* tumorigenicity. Alternatively, CSCs can be cultured in Matrigel™ (Clarke, 2005), an extracellular matrix (ECM)-like mixture primarily made up of laminin and collagen IV that is secreted by the Engelbreth-Holm-Swarm mouse sarcoma cell line. Since Matrigel™ more closely mirrors the *in vivo* environment experienced by cancer cells, assessing clonogenicity and anchorage-independent growth in a three-dimensional growth matrix like Matrigel™ is likely to be more reflective of the true tumorigenic nature of cultured cells versus soft agar or two-dimensional surface culture. Myofibroblasts or other tumor-associated cells can also be mixed into the culture matrix to assess their effects on CSC behavior.

To conduct multiple experiments on isolated CSCs, *in vitro* culture is an option to generate the requisite number of cells. Since normal culture medium often contains fetal calf serum, the heterogeneous mixture of growth factors could trigger the differentiation of the CSCs (Vermeulen, et al., 2010). To protect against such an eventuality, a serum-free medium consisting of a defined set of additives is often used to culture CSCs. Many laboratories use their own proprietary blends of additives, but most defined media for CSC culture contain the growth factors epidermal growth factor (EGF) and basic fibroblast growth factor (bFGF) (Ricci-Vitiani, et al., 2007), as well as the hormones and salts insulin, transferrin and sodium selenate (Kirkland, 2009).

To test the *in vivo* tumorigenicity of CSCs, the gold standard is the animal xenograft. CSCs are mixed with Matrigel™ and injected either heterotopically or orthotopically into immunocompromised mice to assess tumor formation. NOD/SCID mice are the normal animals used, although the activities of B- and T-lymphocytes are not completely ablated. Recently, a new strain of immunocompromised mice, NOD/SCID IL2Rγ^{null}, has been developed that completely ablates lymphocyte activity (Quintana, et al., 2008)—it is thought that these new mice would be less able to oppose tumor formation through xenograft rejection, improving the accuracy of experiments. Although most xenografts are heterotopic (the normal injection site is in the mouse flank), orthotopic xenograft protocols have been developed for certain cancers (such as pancreatic cancer) (Kim, et al., 2009) to provide a microenvironment more like that of the original source of the CSC primary tumor.

1.3.5 Current evidence

1.3.5.1 Prospective identification

The initial studies identifying CSCs in colorectal cancer found CD133 to be a useful marker of cancer stemness. O'Brien and colleagues (2007) found that sorting primary human tumor sample cells on the basis of CD133 expression and injecting $CD133^+$ cells into the renal capsules of NOD/SCID mice (a less heterotopic injection location than the standard subcutaneous injection site) enhanced tumor formation, to the extent that their data indicated a 200-fold enrichment in CSCs. Ricci-Vitiani and others (2007) compared $CD133^-$ cells to $CD133^+$ cells from primary tumor tissues and concluded that the positive population harbored the CSC population and was uniquely able to generate serial mouse xenograft tumors. Todaro and colleagues (2007) confirmed these findings in primary tissue, and Ieta and colleagues (2008) did so in cell lines. What is distinctive about these studies is that both did not rigorously isolate the CSC population—CD133 positivity enriched for CSCs, but did not generate a pure population. This theme of enrichment, not isolation, has been a common one for prospective colorectal CSC identification—it is likely that the CSC compartment must be characterized by numerous factors in order to approach true isolation.

A co-expression study (Kawasaki, et al., 2007) with the colorectal cancer cell lines HT29, HCT116 and SW28 found CD200, a surface glycoprotein involved in innate immune system regulation, to be more than 83% co-expressed with the CD133+ population, while another study found musashi 1, a putative marker of colonic stem cells, to be highly-expressed in primary tissue $CD133^+$ cells (Todaro, et al., 2008). However, when patient metastatic colorectal cancer biopsies were analyzed (Shmelkov, et al., 2008), $CD133^-$ cells were found to be just as tumorigenic as $CD133^+$ cells, so the value of CD133 as a

marker of cancer stem cells may still require interrogation. Indeed, Kemper and colleagues (2010) have recently reported that serum-differentiated primary tissue $CD133^+$ colorectal CSCs only lose expression of the antibody-recognized surface epitope, AC133, not the actual expression of the CD133 protein or transcriptional activity at the CD133 promoter.

Besides CD133, there are a multitude of additional antigens that have the potential to be markers of colorectal CSCs. Dalerba and colleagues (2007) found a rare population of $EpCAM^{high}CD44^+$ cells in six primary colorectal tumor tissues that possessed the ability to initiate NOD/SCID mouse xenograft tumors. The molecule CD44, although an important cell surface receptor for hyaluronic acid that is expressed ubiquitously on colorectal epithelial cells, has been associated with colorectal CSCs in primary tissue (Chu, et al., 2009), and siRNA knockdown of CD44 severely dampens clonogenicity and tumorigenicity (Du, et al., 2008), indicating its functional relevance to colorectal carcinogenesis. Cells co-expressing high levels of CD24 and CD44 ($CD24^{high}CD44^{high}$) have also been associated with the CSC phenotype in colorectal cancer cell lines (Yeung, et al., 2010). CD44 experiments revealed that expression of the intracellular enzyme aldehyde dehydrogenase 1 (ALDH1) is co-localized in the $CD44^+$ cell population (Dylla, et al., 2008) in mouse xenografts, indicating ALDH1 as another possible CSC marker. Experiments that have isolated $ALDH1^+$ colorectal cancer cells from patient biopsies confirmed this hypothesis (Huang, et al., 2009). The putative intestinal stem cell markers musashi 1 and ephrin B2 receptor were also found to be co-expressed in primary $CD44^+$ CSCs (Schulenburg, et al., 2007). Suspension spheroid culture of biopsied colorectal cancer cells, which is also thought to enrich for CSCs, isolated a population of cells expressing CD24, CD29 (also known as $\alpha1\beta2$-integrin) and the aforementioned small intestinal stem cell marker LGR5, among other suspected colorectal CSC markers

(Vermeulen, et al., 2008). Indeed, Kirkland and Ying (2008) found that CD29-expressing cells from the HRA19 colorectal cancer cell line were able to differentiate *in vitro* into columnar, goblet and enteroendocrine cells, suggesting a role for CD29 in stem cell-associated differentiation pathways. Finally, xenografted $CD26^+$ primary bowel cancer cells were found to possess metastatic capabilities in immunocompromised mice, while $CD26^-$ cells were not (Pang, et al., 2010).

Most of the markers associated with colorectal CSCs, namely ALDH1, CD24, CD29, CD44 and CD133, have also been associated with cancer stemness in other solid tumors. As stated previously, breast cancer CSCs were originally identified using the markers CD24 and CD44 (Al-Hajj, et al., 2003), while glioblastoma CSCs were purified using CD133 (Singh, et al, 2004). The CSC population in gastric cancer has recently been determined to be $CD44^+$ (Takaishi, et al., 2009). CSCs derived from pancreatic tumors, long known for its resistance to most standard therapies, have been labeled using CD24, CD44 and EpCAM antibodies (Li, et al., 2009), and primary hepatic CSCs have been found to express the epithelial lineage marker EpCAM (Yamashita, et al., 2009).

1.3.5.2 Molecular biology

Although multiple prima facie biomarkers have been identified that purport to label the CSC compartment in a variety of solid tumors, including colorectal cancer, their functional relevance has not been interrogated completely. It is possible that no functional significance for some of them, such as CD133 (Horst, et al., 2009), will be found. This poses a problem from two perspectives, therapeutic and molecular biologic. Therapeutically, it would be necessary for a new drug to target a particular protein that, if inhibited, would selectively eliminate the CSCs. Molecularly, it would be necessary to identify functional pathways and pathway elements to understand more fully the

mechanisms that drive CSC behavior. As a result, numerous studies have analyzed CSCs for activation (or inhibition) of key cell signaling pathways and transcription factors. Four major pathways will be highlighted here, namely the Polycomb group, Hedgehog signaling, the Wnt pathway and the Notch pathway.

The Polycomb group is an assemblage of proteins that, in concert, act as a transcription modifier. Consisting of two repressor groups, Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2), the Polycomb proteins assist in the modification of chromatin structure through histone methylation and deacetylation to repress the transcription of target genes, such as the Homeobox development genes (Gil, et al., 2005). Within the PRC1 and PRC2 aggregates are two key proteins, BMI1 and EZH2. BMI1, a member of PRC1, primarily inhibits the *Ink4a*/*Arf* gene locus, the proteins $p16^{Ink4a}$ and $p19^{Arf}$, which are known to be important in the activation of the Rb and p53 tumor suppressors. Overexpression of BMI1 leads to an extension of cellular lifespan, while BMI1-null cells senesce quickly—as a result, BMI1 has been found to be important in the maintenance of hematopoietic, neural and pancreatic stem cells, implying a strong role for it in CSC maintenance (Lobo, et al., 2007). Indeed, BMI1 expression has been identified in crypt base cells within the mouse small intestine (Sangiorgi, et al., 2008). In the same vein, EZH2, a member of PRC2, has been found to be overexpressed in poorly differentiated and metastatic tumors (Gil, et al., 2005). These observations imply a role for Polycomb group activity in CSC biology, an implication borne out by experimental results. The head and neck squamous cell carcinoma cell line SASVO3 has been shown to exhibit CSC characteristics, including mouse xenograft tumorigenicity and to express elevated levels of BMI1 (C. Y. Chen, et al., 2009), while in primary glioblastoma tumors, BMI1 is highly expressed in the $CD133^+$ CSC compartment (Abdouh, et al., 2009). Furthermore, when BMI1 expression is ablated, CSC

clonogenicity is inhibited (Abdouh, et al., 2009). In $CD44^{+}EpCAM^{-}$ bladder tissue CSCs, expression of BMI1 and EZH2 are high compared to the non-CSC population (Yang and Chang, 2008).

The Hedgehog signaling pathway is linked to the Polycomb group proteins; indeed, BMI1 is a target of Hedgehog signaling (Lobo, et al., 2007). Pathway components include the secreted Sonic hedgehog (SHH) extracellular ligand, the cell receptor Patched (PTCH) and the transmembrane effector Smoothened (SMO) (Medina, et al., 2009). Expression of SHH is correlated with the regulation of stem cell self-renewal and proliferation, especially with regards to stem cell differentiation (Medina, et al., 2009). SHH is especially important in intestinal development—a lack of SHH expression is associated with the absence of colonic crypts (Madison, et al., 2005). Within colorectal cancer, aberrant SHH signaling is thought to be associated with the progression of tumors from adenoma to carcinoma (Yoshikawa, et al., 2009), implying a role for SHH in maintaining cell tumorigenicity. As a result, it comes as no surprise that aberrant Hedgehog signaling is implicated in CSC biology. Inhibition of SHH in neuroblastoma cell lines abrogates cell tumorigenicity, and siRNA knockdown of the GLI2 Hedgehog target protein ablates clonogenicity (Mao, et al., 2009). In addition, primary pancreatic CSCs have been shown to express high levels of SHH (Lee, et al., 2008), while among $CD133^{+}$ primary tissue bowel CSCs, active Hedgehog signaling is required for cells to survive and initiate xenograft tumors in mice (Varnat, et al., 2009).

The SHH ligand also interacts with another pathway, the Wnt pathway. In its canonical organization, a secreted Wnt ligand (WNT1 to WNT16) binds to the Frizzled membrane receptor. Upon binding, the axin/GSK3β/APC protein complex disassembles, allowing the cytoplasmic transcription factor β-catenin to translocate to the nucleus to promote the

transcription of target genes (Espada, et al., 2009), including cyclin D1, c-Myc and Etv4, all genes associated with proliferation. In the intestine, ablation of β-catenin induces small intestinal cell differentiation in mice (Fevr, et al., 2007), implying a role for the Wnt pathway in intestinal stem cell maintenance. Furthermore, crypt base cells express higher levels of Wnt ligands, such as WNT3 and WNT6, than do crypt luminal cells, again implying a role for the Wnt pathway in regulating intestinal stem cells (Gregorieff, et al., 2005). Within the CSC realm, activation of Wnt signaling has been associated with expansion of the $CD44^+$ CSC population in gastric cancer (Ishimoto, et al., 2010), while high expression of Wnt pathway components has been found in an empirically-defined CSC compartment in urothelial tumors (He, et al., 2009). Stimulation of the Wnt pathway in prostate CSCs promotes cellular self-renewal and triggered increased expression of CD44 and CD133 (Bisson, et al., 2009). Within the large bowel, the Wnt pathway has been found to be preferentially expressed in the CSC compartment of primary tissue cells through the use of a GFP reporter (Vermeulen, et al., 2010).

The Notch pathway is a fourth major regulatory pathway implicated in CSC regulation. Consisting of an extracellular ligand (generally Delta or a member of the Jagged or Serrate protein families) binding to the Notch membrane receptor, which triggers the cleavage of the Notch intracellular domain (NICD), the Notch pathway alters the expression of a multitude of target genes when NICD binds to the DNA binding protein CSL (Bolos, et al., 2009). These target genes include the Hes family of development and patterning genes, c-Myc, Skip2, p27KIP1 and cyclin D1. As a result, Notch signaling is critical for proper tissue development and proliferation—lack of Notch signaling prevents the proper differentiation of cells in the heart, blood, skin, stomach and prostate (Bolos, et al., 2009). Besides its crucial development role, the Notch receptor is overexpressed in a

variety of cancers, including breast and colorectal cancers. In glioma primary cells, pharmaceutical inhibition of the Notch pathway sensitizes CSCs to radiation, while constitutive activation of the pathway promotes radioresistance (Wang, et al., 2010). Furthermore, an empirically identified CSC population found in ovarian tumors was found to overexpress Notch-1 versus the non-CSC population (Zhang, et al., 2008). As a result of these observations, and many others, one of the first clinical trials devoted to developing a drug targeting CSCs is an anti-DLL4 monoclonal antibody, an antibody targeting a Notch pathway ligand (Hoey, et al., 2009).

Clearly, the Polycomb group, Hedgehog pathway, Wnt pathway and Notch pathway are all important in the normal functioning of stem cells and, when dysregulated, in tumorigenesis. In order to understand CSCs fully, the roles of these pathways in maintaining the CSC phenotype and tumor-initiating abilities must be, and have been, examined.

1.3.5.3 Relationship with fibroblasts

As previously stated, the role of stromal cells, such as fibroblasts or myofibroblasts, is critical in creating the proper microenvironment for normal tissue stem cell functioning and proliferation. As a result, it is likely that fibroblasts are critical for proper CSC functioning. Since most experiments to date with CSCs have involved culturing them in isolation, it is possible that co-culturing them with fibroblasts may enhance their ability to self-renew and initiate tumors.

In colorectal cancer, adjacent fibroblasts have been found to promote the migration and invasion of cancer cells. In addition, such fibroblasts are thought to modulate the production of Wnt ligands, thus promoting the nuclear localization of β-catenin in nearby

epithelial cancer cells (Le, et al, 2008). In breast and prostate cancer, secretion of SDF-1 (stromal cell-derived factor-1) by neighboring fibroblasts promotes angiogenesis, while in mice, enriching cancer cell xenografts with fibroblasts promoted their resistance to anti-angiogenesis therapy (Shimoda, et al., 2010). Co-culture experiments with fibroblasts support these findings. Experiments with a murine 4T1 breast cancer model found that elimination of cancer-associated fibroblasts dampened angiogenesis, suppressed metastasis and increased the chemosensitivity of the tumors to doxorubicin (Liao, et al., 2009). The role of fibroblasts in mediating chemoresistance was confirmed in lung cancer, in which *in vitro* co-culture of cancer cells with fibroblasts enhanced resistance to the tyrosine kinase inhibitors gefitinib and erlotinib (Wang, et al., 2009). Even more telling, in oral squamous cell carcinoma, the growth of cell lines in a fibroblast-containing matrix empirically isolated cells capable of initiating cancer cell colonies, a behavior associated with CSCs (Mackenzie, 2004). Finally, factors secreted by fibroblasts, such as interleukin-8 (IL-8) have been confirmed to promote invasiveness in the colorectal cancer cell lines Colo678 and DLD1 (Mueller, et al., 2007).

Clearly, fibroblasts influence CSC behavior and molecular biology—the extent of these influences remains a topic of scientific investigation.

1.4 Control of stemness

1.4.1 CD55

Within the context of cancer stemness, the protein CD55, and its associated gene *CD55*, was originally highlighted in 2007 by Xu and colleagues (2007) when they found that a subpopulation of breast cancer cells, termed the “side population,” uniquely co-expressed

CD55. Side population cells, cells able to efflux the DNA intercalating dye Hoechst 33342, were thought to harbor the CSC population; thus, the observed co-expression suggested that CD55 was a potential prospective marker of cancer stemness. A follow-up study (Ikeda, et al., 2008) focusing specifically on cell separation on the basis of CD55 expression in the MCF7 breast cancer cell line confirmed this suggestion—$CD55^+$ cells were far more tumorigenic than $CD55^-$ cells in murine xenografts. In addition, clinical comparisons of CD55-high breast tumor biopsies with CD55-low biopsies revealed a statistically significant correlation between CD55 expression and poor prognosis; the same observation has been made for colorectal tumors (Durrant, et al., 2003). Combined with observations that CD55 is overexpressed in a variety of cancer environments, especially in colorectal cancer, compared to normal tissues (Li, et al., 2001), extant findings suggest that CD55 could be a useful molecule to study within the context of CSCs.

The gene *CD55*, also known as decay-accelerating factor (*daf*) is a 40 kb gene located on chromosome 1 at locus 1q32 (Mikesch, et al., 2006). Containing eleven exons, *CD55* possesses two splice variants, a 2.2 kb and a 1.5 kb transcript, which correspond to the two CD55 protein isoforms of weight 50 to 100 kDa, a transmembrane receptor and a secreted factor (Mikesch, et al., 2006). The more common receptor form consists of four 60-amino-acid short consensus repeat (SCR) regulatory regions linked to a glycosyl phosphatidyl inositol (GPI) membrane anchor. These SCRs are important in the main function of CD55, the inhibition of the immunological complement pathway. By favoring the inactivation of the C3/C5-convertase enzyme, CD55 acts to prevent the death of the cell attributable to complement activation (Mikesch, et al., 2006).

The ability of CD55 to suppress the canonical complement pathway has important implications for antibody-based cancer therapy. Because antibodies can induce the production of complement proteins to eliminate cells, a process known as complement-dependent cytotoxicity (CDC) (Gancz and Fishelson, 2009), any complement alterations could affect the efficiency of endogenous and therapeutic antibodies. Indeed, membrane expression or secretion of complement regulators, including CD46, CD55 and CD59, have been associated with inhibition of the complement pathway (Gancz and Fishelson, 2009). As a result, drug-resistant cancer cells and cancer cells resistant to tumor-specific antibodies in *in vitro* models have been found to express high levels of complement regulators. Furthermore, CD55 directly affects tumor growth. Stable transfection of an siRNA construct ablating CD55 expression in PC-3 prostate cancer cells (Loberg, et al., 2006) dampens tumor growth in mouse xenografts, while *in vitro* experiments with complement-mediated lysis showed that active CD55 on PC-3 and DU145 prostate cancer cells inhibited cancer cell lysis (Loberg, et al., 2006).

As a result of its roles in resisting antibody drug treatments and in promoting tumorigenicity, CD55 is a molecule of interest in understanding factors that stimulate or maintain colorectal cancer stemness.

1.4.2 BMI1

As previously discussed, the Polycomb group of proteins, of which the protein BMI1 is a member, form an important regulatory process for directing the development of tissues and the behavior of multipotent stem cells, implying a role for the protein, and its associated *BMI1* gene, in the control of CSCs. Its role in colorectal carcinogenesis has already been proven; not only is it overexpressed in an array of solid tumors, including colorectal cancer (Kim, et al., 2004), but BMI1 expression has been correlated with poor

survival in colorectal cancer (Du, et al., 2010). Moreover, the dysregulation of BMI1 expression seems to be correlated with the multistep pathway toward a colorectal tumor--BMI1 expression rises as low-grade colonic lesions grow into carcinomas (Tateishi, et al., 2006).

The 10 kb-long *BMI1* gene (B lymphoma Mo-MLV insertion region 1), located on chromosome 10 at locus 10p13, encodes a 326-amino acid protein using a ten-exon mRNA transcript of weight 46 kDa (Alkema, et al., 1993). The key protein structural landmark within BMI1 is the RING finger motif, a cysteine-rich zinc ion binding domain that is critical for protein-protein interactions involving BMI1 (Hemenway, et al., 1998), such as those with the other members of the PRC1 regulatory complex. As stated earlier, PRC1, of which BMI1 is a constituent along with RING1B, PH1 and CBX4, acts to silence target genes (such as Hox genes) to maintain a stem-like cell phenotype through the blocking of the binding of transcription factors and the ubiquitinylation of histone H2A (Simon and Kingston, 2009). Specific targets of BMI1-mediated silencing include the proteins $p16^{Ink4a}$ and $p19^{Arf}$ (Gil, et al., 2005). This maintenance of a stem-like phenotype has been observed in fibroblasts, in which forced expression of BMI1 dampens $p16^{Ink4a}$ and $p19^{Arf}$ expression and induces the immortalization of the fibroblasts (Jacobs, et al., 1999).

The role BMI1 plays in CSC maintenance has been well-studied. In colorectal cancer, experiments with cell lines have revealed an association between a CSC-like behavior, such as increased clonogenicity and downregulation of differentiation-associated proteins expression, and a promotion of BMI1 expression (Kirkland, 2009). SiRNA knockdown of BMI1 expression in tumor-isolated neuroblastoma cells induced them to undergo apoptosis very quickly, while knockdown in non-cancerous cells had little effect (Liu, et

al., 2006). BMI1 is highly expressed in primary breast cancer cells cultured as nonadherent spheroids (Liu, et al., 2006), and mouse xenografts of these BMI1-expressing cells produced large tumors, while mouse xenografts of cells with low-to-moderate BMI expression produced much smaller tumors. Indeed, xenografts generated from cells with siRNA-reduced BMI expression did not grow (Liu, et al., 2006). CSC populations of $CD44^+$ primary head and neck squamous cell carcinoma cells express high levels of BMI1 (Prince, et al., 2007), while the highly-malignant I-type neuroblastoma cells (which possess the CSC behavioral characteristics of high clonogenicity and self-renewal) express BMI1, while more differentiated N-type cells express less BMI1 (Cui, et al., 2006).

As a result of its roles in promoting the stem-like phenotype and in promoting tumor initiation, BMI1 is an item of interest in understanding factors that maintain colorectal cancer stemness.

1.4.3 Other factors

1.4.3.1 CD24

The CD24 protein was originally identified as a B-cell glycosylated membrane antigen in mice, but its normal human epithelial expression is quite limited (Lim, 2005). In addition, CD24 is highly expressed in a variety of solid tumors (Lim, 2005), in addition to its aforementioned presence on CSC populations. The associated *CD24* gene, located on Chromosome 6 at locus 6q21, is only 2.2 kb in length, yielding a small protein of 80 amino acids of weight 8 kDa. The membrane-bound GPI anchor of CD24 is linked to a glycosylated extracellular protein core, which mediates CD24 functions. There is circumstantial evidence that CD24 is involved in direct regulation of membrane tyrosine

kinase activation, but its more characterized function involves the regulation of intercellular interactions through the binding of P-selectin, an endothelial protein that alters the binding properties of the glycosyl groups on CD24 (Lim, 2005), although the specific roles of CD24 in cellular functions (particularly in oncogenic functions), if any, have not been fully-characterized.

1.4.3.2 CD44

As opposed to CD24, the CD44 protein is a much larger molecule. Located on Chromosome 11 at locus 11p13, the 93.5 kb CD44 gene generates an mRNA transcript from 20 exons, 12 of which are found in the most common transcript. Eleven additional transcripts are present (CD44v1 to CD44v11) and are defined by the insertion of one of eleven additional exons. A generated 742-amino acid protein of weight 81.5 kDa is a transmembrane protein, consisting of extracellular domain (the exon insertions affect the membrane-proximal domain segment), a single transmembrane domain and an intracellular domain (which can also vary in length). The extracellular domain can also be glycosylated or glycosaminoglycolated (Cichy and Puré, 2003). The functions of CD44, which is broadly expressed on many cell types, also include the regulation of cell-to-cell and cell-to-matrix interactions via the CD44 extracellular domain. This domain mainly binds the extracellular matrix components hyaluronan and collagen, but it also interacts with the ligands osteopontin, matrix metalloproteinases (MMPs) and ERM proteins (such as villin-2 and radixin) (Cichy and Puré, 2003).

1.5 Control of differentiation and CDX1

Within the context of intestinal differentiation, CDX1, or caudal-type homeobox protein 1, is one of the most important transcription factors. An 18 kb-long gene at locus 5q33.1 on Chromosome 5, *CDX1* gives rise to two alternatively-spliced transcripts that both generate a 265-amino acid protein of weight 28 kDa. This protein shares structural similarities to the two other members of the caudal-related protein family, CDX2 and CDX4, in the central homeodomain region—two other characterized regions include a cytoplasm export signal and a transactivation domain (Guo, et al., 2004).

CDX1 mainly promotes the proper patterning and differentiation of the small and large intestine. Early in murine embryo development, for example, CDX1 expression emerges in the developing intestinal epithelium by 12.5 days to direct its development (Guo, et al., 2004), although CDX1-null mice possess normal intestines (Bonhomme, et al., 2008). In the adult human, overall CDX1 expression increases along the length of the large intestine, while within colonic crypts, CDX1 expression is lowest at the base and highest toward the luminal surface (Chan, et al., 2009), implying, as stated earlier, that CDX1 is a differentiation factor. Its mechanism of action consists of binding to CDX-responsive elements (with consensus sequence TTTATA/G or TTTAT/C) in the promoter sequences of target genes to trigger gene transcription. These target genes include MUC2, KLF4 and GPA33, among many others (Guo, et al., 2004).

Its expression and effects in colorectal cancer have also been studied extensively. CDX1 expression has been found to be downregulated in tumors compared to normal tissue (Pilozzi, et al., 2004), and that downregulation can be attributable to methylation of the CDX1 promoter (Chan, et al., 2009). Expression of CDX1 in the Colo205 cell line induced a differentiated phenotype (Keller, et al., 2004), while induced CDX1 expression

reduces the expression of cyclin D1, a gene associated with cell proliferation (Lynch, et al., 2003). Thus, it is clear that CDX1 is associated with the promotion of differentiation—perhaps the lack of CDX1 expression in a compartment of cancer cells could highlight the CSC population.

1.6 Summary

1.6.1 Aims

As a result of the increased interest in CSC biology, particularly in solid tumors such as colorectal cancer, it is clear that more work is necessary to understand the process of differentiation and stemness in colorectal cancer. By using a colorectal cancer cell line panel, the experiments described herein attempt to investigate the following phenomena:

1. Identify cancer cell lines that possess the ability to spontaneously differentiate *in vitro*;
2. Identify differentially-expressed genes associated with spontaneous differentiation or lack thereof;
3. Ascertain the usefulness of CD55 as a prospective marker of colorectal cancer stem cells, both alone and in concert with the markers CD24 and CD44;
4. Manipulate the expression of differentially-expressed genes by cDNA induction to ascertain their importance in cancer stem and differentiation behavior;
5. Introduce promoter-driven fluorescent plasmid constructs into cell lines to identify the usefulness of CDX1 as a negative marker of cancer stemness; and

6. Co-culture colonic myofibroblasts with cell lines to identify their effects on cancer stem cell populations and behavior.

1.6.2 Cell line model

The cancerous materials mainly used to make the aforementioned observations were cell lines, cells cultured and serially-passaged *in vitro* for years from resected primary and metastatic colorectal tumors. While there are many questions about the usefulness of such cells in understanding cancer, which is a disease often dependent on extracellular factors provided *in vivo*, a meta-analysis has found cell lines to be a useful model for understanding single nucleotide polymorphisms in neuroblastoma because of inevitable stromal contamination of primary tissues (Volchenboum, et al., 2009). An analysis of copy number variations in colorectal cancer also found a high degree of similarity between primary tumors and associated derived cell lines (Douglas, et al., 2004), while epigenetic assessments of cell line pairs, derived from the same tumor but cultured separately for years, have found broad similarities between the lines (Dikomitou, 2009). Clearly, cell lines can serve as a useful model for studying cancer biology and can do so while providing flexibility in terms of treatment conditions and gross cancerous material available for experimentation.

CHAPTER 2

MATERIALS AND METHODS

2.1 Reagent and goods suppliers

Unless otherwise specified, inorganic buffers, fine chemicals and DNA primers used were from Sigma-Aldrich (Poole, UK). Tissue culture flasks, plates, tubes and other experimental consumables were from Corning (Corning, New York, USA). Tissue culture biologics, including media, antibiotics and nutrient supplements, were from Invitrogen (Paisley, UK). Purified proteins, including growth factors, were from Peprotech (London, UK). Restriction enzymes, buffers and ligation reagents were from New England Biolabs (Ipswich, Massachusetts, USA).

When pre-assembled kits of reagents were used, all referenced reagents were from the named kit manufacturer.

2.2 Two-dimensional cell culture

2.2.1 Cell lines

The human colorectal cancer cell lines C80, C84, C99, CC20, CCK81, Colo741, DLD1, Gp2D, HCA7, HCT116, HCT15, HRA19, HT29, LIM1863, Lovo, LS174T, LS180, LS411, LS513, NCIH716, PCJW, RCM1, RKO, SKCO1, SNUC2B, SW1222, SW1417, SW48, Vaco5 and WIDR were obtained from the cryogenic storage of the Cancer and Immunogenetics Laboratory (CIL), Weatherall Institute of Molecular Medicine (WIMM), Oxford, UK. Characterizations of their origins and mutational statuses can be found in Appendix A.

The C80, C84 and C99 cell lines were originally established in the CIL by David Bicknell. The SW1222 cell line was a gift from Meenhard Herlyn of the Wistar Institute,

Philadelphia, Pennsylvania, USA. The remaining cell lines were originally obtained from the American Type Culture Collection (ATCC) (Manassas, Virginia, USA), German Collection of Microorganisms and Cell Cultures (DSMZ) (Braunschweig, Germany) or European Collection of Cell Cultures (ECACC) (Salisbury, UK).

The human colonic myofibroblast cell line CCD-18Co was purchased from the ATCC. CCD-18Co has previously been characterized to be a myofibroblast cell line through analysis of its growth, morphology and protein expression (Valentich, et al., 1997).

2.2.2 Culture conditions

The cell lines C80, C84, C99, HRA19, RCM1 and Vaco5 were cultured in Iscove's modification of Dulbecco's Modified Eagle's Medium (IMDM). The cell lines Colo741, LS513, NCIH716, RKO and SNUC2B were cultured in Roswell Park Memorial Institute (RPMI) 1640. All remaining cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM), including CCD-18Co. The cell lines LIM1863, NCIH716 and PCJW grew in suspension; all others were adherent.

For routine cell culture and for most experiments, media were supplemented with 10% fetal bovine serum (FBS) (Biosource, Nivelle, Belgium), 1% L-glutamine and 1% penicillin/streptomycin to create complete media. For serum-free culture experiments, media were supplemented with insulin-transferrin-selenium-X supplement mixture (Kirkland, 2009), 10 nanograms/milliliter (ng/mL) epidermal growth factor (EGF) and 20 ng/mL basic fibroblast growth factor (bFGF) (Ricci-Vitiani, et al., 2007). All cells (adherent and suspension) were routinely grown in T75 (75 cm^2 growth surface) vented tissue-culture-treated flasks with 10 mL of medium.

DMEM- and IMDM-cultured lines were incubated at 37 ºC in a 10% CO_2 Heracell 150 humidified chamber (Thermo Scientific, Waltham, Massachusetts, USA), while RPMI lines were similarly incubated at 5% CO_2.

2.2.3 Cell culture maintenance

Once DMEM- and RPMI-cultured adherent cells reached approximately 70-80% confluence, cells were passaged into new flasks by washing cells with 10 mL absolute phosphate buffered saline (PBSA) and detached by incubating cells with 2 mL trypsin/EDTA solution for 5 min. Trypsin was inactivated by adding 10 mL medium. Subsequently, cells were harvested by centrifugation at 1,500 rpm for 5 minutes (min) and resuspended in 10 mL medium for subculturing. Based on cell line growth characteristics, cell suspensions were divided into new flasks at a 1:5 to 1:10 dilution.

Once IMDM-cultured adherent cells reached approximately 70-80% confluence, cells were passaged into new flasks by washing cells with 10 mL calcium-free Suspension Minimal Essential Medium (SMEM) and detached by calcium depletion by incubating cells with 10 mL SMEM overnight in a humidified incubator. Subsequently, cells were harvested by centrifugation at 1,500 rpm for 5 min and resuspended in 10 mL IMDM for subculturing. Cell suspensions were divided into new flasks at a 1:10 dilution.

For cell lines grown in suspension, cells were harvested by centrifugation at 1,500 rpm for 5 min and resuspended in 10 mL medium for subculturing. Cell suspensions were divided into new flasks at a 1:10 dilution.

Cells were routinely counted on an improved Neubauer hemocytometer (Hawksley, Lancing, UK). An aliquot of 100 microliters (µL) of cell suspension was stained with 100 µL trypan blue dye to allow for exclusion of dead cells from the count. Subsequently, 10

µL of the stained cell suspension was spotted onto the hemocytometer and counted on a BH2 light microscope (Olympus, Southend-on-Sea, UK) to calculate density.

2.2.4 *Mycoplasma* testing

Cell cultures were routinely tested for *Mycoplasma* contamination using the MycoAlert® Mycoplasma Detection Assay (Lonza, Rockland, Maine, USA). Briefly, after two days of cell culture, medium was withdrawn and centrifuged at 1,500 rpm for 5 min to pellet cell debris. Of the cleared supernatant, 100 µL was transferred to one well of a white-walled 96-well plate, combined with 100 µL MycoAlert® Reagent and incubated for 5 min at room temperature. The luminescence of the well was read on a FLUOstar Optima luminometer (BMG Labtech, Offenburg, Germany) ("Reading A"). Subsequently, 100 µL MycoAlert® Substrate was added to the well and incubated for 10 min, after which the luminescence was read again ("Reading B").

The ratio of Reading B to Reading A was computed; a ratio greater than 2 indicated Mycoplasma contamination, while a ratio less than 1 indicated no contamination. If the value were between 1 and 2, the cell culture would be retested. Any contaminated cell cultures were immediately discarded.

2.2.5 Cell culture storage and retrieval

Cell cultures were routinely frozen and stored in a cryogenic freezer within the vapor phase of liquid nitrogen at -150 ºC. Cells were detached or harvested and centrifuged at 1,500 rpm for 5 min. The resulting cell pellet was resuspended in 1 mL cryogenic freezing medium (10% dimethyl sulfoxide in FBS), aliquotted into a cryogenic vial and frozen for 1 hour (hr) on dry ice. Resulting frozen cells were placed in a CBX 40 cryogenic freezer (Wessington Cryogenics, Houghton le Spring, UK) for storage.

2.3 Three-dimensional cell culture

Cell cultures were grown in various three-dimensional matrices, namely growth-factor-reduced Matrigel™ (BD Biosciences, Abingdon, UK) or a 1:1 mixture of Matrigel™ and collagen I (Inamed, Santa Barbara, California, USA) (for initial colony morphology assessments on a cell line panel). Matrigel™, a mixture of basement membrane proteins (including laminin and collagen IV) secreted by the Engelbreth-Holm-Swarm (EHS) mouse sarcoma cell line, was thought to mimic the constituents of the extracellular matrix (ECM) (Kleinman and Martin, 2005). Adding collagen I stiffened the resulting gelatinous matrix and made three-dimensional cell cultures easier to manipulate.

Collagen I growth solution was prepared from stock by combining 1 mL 10X PBSA and 1 mL sterile water with 8 mL collagen I. Solution pH was adjusted to 7.4 with 5 *M* sodium hydroxide and sterilized using a 0.22-µm filter (Sartorius Stedim, Aubagne, France).

Cells were detached, resuspended in PBSA and diluted 10:1 in Matrigel™ or in a 1:1 mixture of Matrigel™ and collagen I and plated in 1-mL aliquots into wells of a 48-well plate or in 100-µL aliquots into wells of a 96-well plate to yield a final cell count of 1,000 per well. Plates were incubated at 37 ºC in a humidified chamber for 15 min to solidify Matrigel™/collagen I gels. Finally, gels were overlaid with media and incubated for two weeks.

After two weeks, gels were imaged on an Axiovert 135 inverted light microscope (Carl Zeiss, Gottingen, Germany) under phase contrast. Images were photographed and analyzed using the integrated Axiovision image processing software.

For downstream immunohistochemical analysis, gels were washed twice with PBSA, fixed in fixing buffer (1% formalin in PBSA) for 5 min and embedded in 3% agarose (Invitrogen). Subsequent paraffin embedding and slide sectioning was performed by the Department of Histopathology, John Radcliffe Hospital, Oxford, UK.

2.4 Cell growth analyses

2.4.1 Clonogenicity assays

Cells were filtered through a CellTrics® 20-μm filter (Partec, Munich, Germany) to obtain single-cell suspensions. The resulting suspensions were diluted with an appropriate volume of Matrigel™ and plated in 100-μL triplicate aliquots into wells of a 96-well plate to yield a final cell count of 1,000 per well. Plates were incubated at 37 ºC in a humidified chamber for 15 min to solidify Matrigel™/collagen I gels. Finally, gels were overlaid with media and incubated for two weeks. After two weeks, gels were imaged on an Axiovert 135 inverted light microscope under phase contrast. Images were collected and analyzed using the integrated Axiovision image processing software. All colonies larger than 25 cells were counted and distinctive distributions in colony morphology were recorded. Since the cell lines tested all grew at different speeds, colonies larger than 25 (midway between four and five doublings, 16 and 32 cells) were counted such that the counts did not discriminate against slow-growing cell lines. Selecting a higher threshold, such as 32, could have biased counts toward fast-growing cell lines that could proliferate quickly within the short growth span of two weeks.

For assays involving the co-culture of colorectal cancer epithelial cells and CCD-18Co colonic myofibroblasts, single-cell suspensions were combined, diluted with appropriate

volumes of Matrigel™ and plated in 100-µL triplicate aliquots into wells of a 96-well plate to yield a final count of 1,000 epithelial cells and 500, 1,000 or 2,000 myofibroblasts per well.

2.4.2 Proliferation assays

To assess the proliferative capacity of cells, the CellTiter 96® AQ_{ueous} (Promega, Madison, Wisconsin, USA) MTS assay was used. The MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) assay measured cell metabolic activity by photometrically monitoring the *in vitro* reduction of MTS to a formazan product by cellular enzymes. Since only viable cells can catalyze the MTS reduction, the assay excludes dead cells from measurement (Promega, 2009).

Briefly, cell suspensions were plated in triplicate into wells of a 96-well plate to yield a final cell count of 3,000 per well and overlaid with media. After incubating plate for 1-4 days, media were replaced with 120 µL diluted MTS/PMS reagent (20 µL MTS/PMS reagent in 100 µL medium) per well, and plates were incubated for 2 hr at 37 ºC in a humidified chamber. Subsequently, the absorbance of wells at 490 nm was measured on a µQuant Microplate Reader (Biotek Instruments, Winooski, Vermont, USA) and analyzed using the integrated KC4 software.

For assays involving the co-culture of colorectal cancer epithelial cells and CCD-18Co colonic myofibroblasts, cell suspensions were combined and plated in triplicate into wells of a 96-well plate to yield a final count of 3,000 epithelial cells and 3,000 myofibroblasts per well.

2.5 Genetic analyses

2.5.1 General DNA processing methods

2.5.1.1 DNA quantification

DNA was quantified at a wavelength of 260 nm using an ND-1000 spectrophotometer (NanoDrop Products, Wilmington, Delaware, USA). The integrated NanoDrop 1000 software was used to convert absorbance to DNA concentration using an extinction coefficient of 50 ng-cm/ μL. DNA quality was assessed by computing the ratios of absorbances at 260 nm to 280 nm (OD_{260}/OD_{280}) and at 260 nm to 230 nm (OD_{260}/OD_{230}). An OD_{260}/OD_{280} of 1.8 or greater and an OD_{260}/OD_{230} of 1.8 to 2.2 indicated high-quality DNA (Thermo, 2008).

2.5.1.2 Genomic DNA extraction

The DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) was used to extract genomic DNA (gDNA) from cell lines. Briefly, 5 x 10^6 cells were detached and resuspended in 200 μL PBSA. Cells were lysed with 20 μL proteinase K solution and 200 μL Buffer AL and incubated at 56 ºC for 10 min. Liberated gDNA was precipitated with 200 μL absolute ethanol, bound to a silica spin column by centrifugation and washed with 500 μL Buffer AW1 and Buffer AW2. Purified gDNA was eluted into 200 μL Buffer AE and quantified.

2.5.1.3 Gel electrophoresis

DNA constructs and fragments were resolved by agarose gel electrophoresis. Agarose gels of varying percentage (normally 1-2%) were prepared by dissolving an appropriate mass of electrophoresis-grade agarose (Invitrogen) (normally 0.5-1.0 g) and 10 μL SYBR

Safe™ (Invitrogen) DNA stain in 50 mL tris(hydroxymethyl)aminomethan-acetate (tris-acetate) ethylenediaminetetraacetate (EDTA) (TAE) buffer (40 m*M* tris, 20 m*M* acetic acid and 1 m*M* EDTA). The resulting solution was poured into a mold to set, placed in a Mini-Sub Cell GT electrophoresis unit (Bio-Rad Laboratories, Hercules, California, USA) and overlaid with TAE buffer. DNA samples were combined at a ratio of 4 to 1 with GelPilot loading dye (QIAGEN) and loaded onto the gel alongside 1 kilobase (kb) DNA ladder (New England Biolabs). Samples were run for 60 min at a constant 120 V.

Gels were imaged under ultraviolet illumination on a Gel Doc 2000 scanner (Bio-Rad Laboratories) and photographed using the integrated Quantity One software.

2.5.1.4 DNA purification

The QIAQuick Gel Extraction Kit (QIAGEN) was used to purify DNA fragments resolved through gel electrophoresis or prepared from restriction digests. Briefly, the DNA fragment was excised from the agarose gel, dissolved in 3 mass volumes of Buffer QG at 50 ºC for 10 min and combined with 1 mass volume of isopropanol. Alternatively, the restriction digest was combined with 5 volume equivalents of Buffer QG. The resultant DNA in solution was bound to a silica spin column by centrifugation and washed with 500 µL Buffer QG and 750 µL Buffer PE. Purified DNA was eluted into 50 µL water and quantified.

2.5.1.5 Restriction digests and Klenow fill-in reactions

DNA constructs were digested with restriction enzymes for downstream cloning experiments. To cleave DNA, 1 µg DNA was combined with 5 µL of the appropriate restriction buffer, 0.5 µL of 10X BSA solution (when necessary), 2.0 units of the appropriate restriction enzyme(s) and sterile water for a 50 µL reaction. Reactions were

incubated for 1 hr at 37 ºC (25 ºC for the *Sma*I enzyme). Resulting cleaved DNA was purified and quantified.

To create blunt DNA ends from 5'- or 3'- overhangs, 1 µg cleaved DNA was combined with 3 µL REact2 Buffer (Invitrogen), 1 µL each of 0.5 mM dATP, dCTP, dGTP and dTTP (Sigma-Aldrich), 0.5 units of Klenow large fragment of DNA Polymerase I (Invitrogen) and sterile water for a 30 µL reaction. Reactions were incubated for 10 min at room temperature, purified and quantified.

2.5.1.6 Ligation

Cleaved DNA constructs were ligated to produce complete, circular DNA products suitable for propagation and expression. To ligate a plasmid vector with a shorter insert, 100 ng cleaved plasmid vector DNA were combined with an appropriate quantity of insert DNA at an insert:vector molar ratio of 3:1. To this mixture was added 1 µL 10X T4 ligase buffer, 1.0 unit of T4 ligase and sterile water for a 20 µL reaction. Reactions were incubated overnight at 16 ºC, purified and transformed into *Escherichia coli* (*E. coli*) chemically-competent cells.

2.5.1.7 Bacterial transformation

Circular plasmids were transformed into One Shot® TOP10 chemically-competent *E. coli* bacteria (Invitrogen). Briefly, 2 µL of plasmid DNA solution were combined with 50 µL competent cell solution, incubated for 30 min on ice, heat-shocked for 30 seconds (sec) at 42 ºC and incubated further for 2 min on ice. For recovery, transformation was combined with 250 µL SOC medium and shaken for 1 hr at 37 ºC.

Transformed bacterial colonies were isolated by spreading 25 µL transformation culture on an appropriate antibiotic-selective plate and incubated overnight at 37 ºC.

2.5.1.8 Bacterial plasmid extraction

Resultant isolated colonies from transformations were cultured in 3 mL antibiotic-selective lysogeny broth (LB), either LB-ampicillin (50 µg/mL ampicillin) or LB-kanamycin (50 µg/mL kanamycin), and shaken overnight at 37 ºC. For cloning studies involving the propagation of a DNA vector containing a specific DNA insert, a 20 µL aliquot of the 3 mL bacterial culture was combined with 5 µL 5X cracking buffer (25 g sucrose, 2.5 mL 10% sodium dodecyl sulfate, 2.5 mL 10 *M* sodium hydroxide and 50 mg bromophenol blue dye in 50 mL water) to lyse bacteria. Since supercoiled DNA vectors containing inserts resolve faster than supercoiled DNA vectors alone via gel electrophoresis, bacterial lysates were resolved via gel electrophoresis alongside transformation control lysates to identify colonies containing the desired insert (Todd Lamitina, personal communication).

The QIAprep Miniprep Kit (QIAGEN) was used to extract and purify plasmid DNA from small bacterial cultures found to contain the desired DNA vector and insert. Briefly, bacterial cultures were harvested by centrifugation, resuspended in 250 µL Buffer P1 and lysed with 250 µL Buffer P2. Bacterial debris was precipitated with 350 µL Buffer N3 and collected by centrifugation. The resultant plasmid DNA in the supernatant was bound to a silica spin column by centrifugation and washed with 500 µL Buffer PB and 750 µL Buffer PE. Purified DNA was eluted in 50 µL Buffer EB and quantified.

For large-scale plasmid DNA extraction, isolated colonies from transformations were cultured in 3 mL antibiotic-selective LB and shaken for 8 hrs at 37 ºC. A 100 mL antibiotic-selective LB culture was inoculated with 100 µL starter culture and shaken overnight at 37 ºC.

The EndoFree Plasmid Maxi Kit (QIAGEN) was used to extract and purify stock plasmid DNA from large bacterial cultures. Briefly, bacterial cultures were harvested by centrifugation, resuspended in 10 mL Buffer P1 and lysed with 10 mL Buffer P2. Bacterial debris was precipitated with 10 mL Buffer P3 and removed by filtration through a QIAfilter Cartridge. Endotoxins in the resultant flow-through were inactivated with 2.5 mL Buffer ER. Plasmid DNA in the decontaminated flow-through was bound to an equilibrated QIAGEN-tip 500 by filtration and washed with 60 mL Buffer QC. Purified DNA was eluted in 15 mL Buffer QN, precipitated with 10.5 mL isopropanol and pelleted by centrifugation at 15,000 x *g* at 4 ºC for 30 min. Precipitated pellet was washed with 5 mL ethanol and re-pelleted by centrifugation at 15,000 x *g* at 4 ºC for 10 min. Resulting DNA pellet was dried, dissolved in 500 µL Buffer TE and quantified.

2.5.1.9 DNA sequencing

DNA constructs and products were sequenced at the DNA Sequencing Core Facility, WIMM, on an ABI-3730 48-capillary sequencing machine (Applied Biosystems, Foster City, California, USA) with supplied sequencing primers. Resultant sequences and associated chromatograms were analyzed using the Sequencher® software package (Gene Codes Corporation, Ann Arbor, Michigan, USA).

2.5.2 Promoter construct procurement

2.5.2.1 Design of CDX1 *promoter construct*

The upstream promoter regions of the *CDX1* gene from the human (accession number NM_001804) and mouse (NM_009880.3) genomes were sourced from the Genome Browser online genomic database (University of California, Santa Cruz, USA) at www.genome.ucsc.edu. Promoter sequences of varying lengths, from 100 to 3000 bases

upstream of the transcription start site, from human and mouse genomes were aligned to each other using the ClustalW2 online bioinformatics multiple sequence alignment tool (European Bioinformatics Institute, Cambridge, UK) at www.ebi.ac.uk/Tools/clustalw2 to identify key areas of high interspecies conservation. Default alignment settings were used.

2.5.2.2 CDX1 *promoter PCR reactions*

The polymerase chain reaction (PCR) was used to synthesize an 1143-base *CDX1* promoter fragment corresponding to the -1105 to +38 upstream region. The template gDNA was extracted from the C80 cell line. The forward primer was

5'-CCCAAGGGGGCTCCAGCCAT-3' (Tm = 75.3 ºC) and the reverse primer was 5'-CGACGACGAGCAACCGCTCA-3' (Tm = 76.2 ºC). Primers were designed using the Primer-BLAST online primer design tool (National Center for Biotechnological Information, Bethesda, Maryland, USA) at www.ncbi.nlm.nih.gov/tools/primer-blast and evaluated for quality using the NetPrimer online primer analysis tool (PREMIER Biosoft International, Palo Alto, California, USA) at www.premierbiosoft.com/netprimer.

The AccuPrime™ *Pfx* DNA polymerase (Invitrogen) was used to synthesize the promoter fragment. *Pfx* is a proofreading DNA polymerase extracted from the KOD strain of *Thermococcus* bacteria (Invitrogen, 2010). Briefly, 100 ng C80 gDNA were combined with 5 µL 10X AccuPrime™ *Pfx* reaction mix, 1.5 µL of 10 µ*M* forward and reverse primer solutions, 1.25 units of AccuPrime™ *Pfx* DNA polymerase and sterile water for a 50 µL reaction. Reactions were incubated on a G-StormGS4 thermal cycler (Gene Technologies, Braintree, UK) for 36 cycles per the following conditions: 95 ºC for 15 sec (denature), 60 ºC – 70 ºC temperature gradient for 30 sec (primer anneal), 68 ºC for 60

sec (extension). An initial 2 min denature step at 95 ºC and a final 5 min extension step at 68 ºC were included at the beginning and end of the reaction, respectively.

Reactions were gel-purified and quantified. Purified fragments were sequenced using PCR primers to confirm identity.

2.5.2.3 KRT20 *promoter procurement*

A *KRT20* promoter construct (pGL3-FD), consisting of the first 433 bases upstream of the *KRT20* transcription initiation site cloned into the pGL3-Basic vector, was previously constructed in the CIL by Carol Chan and colleagues (2009). Out of the five promoter constructs synthesized (pGL3-FA through pGL3-FE), the 433-base upstream promoter (pGL3-FD) was found to drive the highest luciferase expression and was thus selected.

2.5.2.4 Control promoter procurement

A *TACSTD1* promoter construct (p3gE), consisting of the first 3,400 bases upstream of the *TACSTD1* transcription initiation site cloned into the pEGFP-1 vector, was obtained from Marianne Rots of the University Medical Center Groningen, Netherlands, from previous work by McLaughlin and colleagues (2004). The p3gE plasmid encodes the promoter for the EpCAM protein, a cell surface protein ubiquitously expressed by colorectal epithelial cells, including the vast majority of colorectal cancer cell lines.

2.5.3 Promoter plasmid construct synthesis

2.5.3.1 Plasmid vector description

The pTurboRFP-PRL-dest1 vector (Evrogen, Moscow, Russia) was purchased to monitor cellular activation of the *CDX1* and *KRT20* promoters. Ligated promoter fragments drove expression of a downstream synthetic, humanized red fluorescent protein (RFP) encoding

a PEST amino acid sequence for temporally-specific assessments of RFP expression. Mammalian plasmid synthesis was driven by the SV40 origin of replication; bacterial synthesis was driven by the pUC origin of replication. Kanamycin and neomycin resistance cassettes conferred resistance to kanamycin for bacterial antibiotic selection and to G418 for mammalian stable selection, respectively.

The pZsGreen1.1 vector (Clontech, Mountain View, California, USA) was also purchased to monitor cellular activation of the CDX1 promoter. Ligated promoter fragments drove expression of a downstream, synthetic ZsGreen1 green fluorescent protein. Mammalian plasmid synthesis was driven by the SV40 origin of replication; bacterial synthesis was driven by the pUC origin of replication. Kanamycin and neomycin resistance cassettes conferred resistance to kanamycin for bacterial antibiotic selection and to G418 for mammalian stable selection, respectively.

2.5.3.2 CDX1 *promoter fragment TOPO blunt-end cloning*

The Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) was used to clone the *CDX1* PCR promoter fragment. Briefly, 2 µL purified PCR product were combined with 1 µL salt solution, 1 µL pCR®-Blunt II-TOPO® DNA vector and 2 µL sterile water and incubated at room temperature for 5 min to ligate the PCR product into the vector. Resulting plasmid was transformed into bacteria, extracted and quantified. Purified pCR®-Blunt II-TOPO®-CDX1_Prom was sequenced using the M13 and T7 universal primers to confirm identity and orientation of PCR insert.

2.5.3.3 CDX1 *target vector cloning*

To clone the *CDX1* promoter fragment into the target pTurboRFP-PRL-dest1 vector, 1 µg pCR®-Blunt II-TOPO®-CDX1_Prom was digested with 2.0 units each of *Hin*dIII and

*Eco*RV restriction enzymes, gel-purified and quantified to isolate the *CDX1* promoter fragment. The target vector was prepared by digesting 1 µg pTurboRFP-PRL-dest1 with 2.0 units each of *Hin*dIII and *Sma*I restriction enzymes, followed by purification and quantification. The promoter fragment insert was ligated into the target vector, transformed into bacteria, extracted and quantified. After sequencing the resulting plasmid, pTurboRFP-PRL-dest1-CDX1_Prom, using the 5'-ACGGTTCCTGGCCTTTTGCT-3' forward and 5'-GTTGTTCACGGTGCCCTCCA-3' reverse primers to confirm identity and orientation of the *CDX1* promoter insert, a plasmid DNA stock was synthesized.

To clone the *CDX1* promoter fragment into the target pZsGreen1.1 vector, 1 µg pCR®-Blunt II-TOPO®-CDX1_Prom was digested with 2.0 units each of *Hin*dIII and *Eco*RV restriction enzymes, gel-purified and quantified to isolate the *CDX1* promoter fragment. The target vector was prepared by digesting 1 µg pZsGreen1.1 with 2.0 units each of *Hin*dIII and *Sma*I restriction enzymes, followed by purification and quantification. The promoter fragment insert was ligated into the target vector, transformed into bacteria, extracted and quantified. After sequencing the resulting plasmid, pZsGreen1-1-CDX1_Prom, using the 5'- TGTTCTTTCCTGCGTTATCCCC-3' forward and 5'- CCGTCCACGCAGCCCTCCAT-3' reverse primers to confirm identity and orientation of the CDX1 promoter insert, a plasmid DNA stock was synthesized.

2.5.3.4 KRT20 *target vector cloning*

To clone the *KRT20* promoter fragment into the target pTurboRFP-PRL-dest1 vector, 1 µg pGL3-FD was digested with 2.0 units each of *Hin*dIII and *Sac*I restriction enzymes, gel-purified and quantified to isolate the *KRT20* promoter fragment. The target vector was prepared by digesting 1 µg pTurboRFP-PRL-dest1 with 2.0 units each of *Hin*dIII and

*Sac*I restriction enzymes, followed by purification and quantification. The promoter fragment insert was ligated into the target vector, transformed into bacteria, extracted and quantified. After sequencing the resulting plasmid, pTurboRFP-PRL-dest1-KRT20_Prom, using the previously detailed sequencing primers to confirm identity and orientation of the *KRT20* promoter insert, a plasmid DNA stock was synthesized.

2.5.4 cDNA construct procurement

A *CDX1* cDNA plasmid construct (pRc/CMV-CDX1) was a gift from Jaleh Malakooti of the University of Illinois, Chicago, Illinois, USA. A *BMI1* cDNA plasmid construct (pCMV-XL5-BMI1) was purchased from OriGene (Rockville, Maryland, USA).

2.5.5 cDNA plasmid construct synthesis

2.5.5.1 Plasmid vector description

pBI-CMV3 (Clontech, Mountain View, California) was purchased to drive the bidirectional expression of the *CDX1* and *BMI1* cDNAs alongside a reporter ZsGreen1 green fluorescent protein. A CMV promoter drove expression of ligated cDNA fragments.

Mammalian plasmid synthesis was driven by the SV40 origin of replication; bacterial synthesis was driven by the ColE1 origin of replication. An ampicillin resistance cassette conferred resistance to ampicillin for bacterial antibiotic selection.

2.5.5.2 CDX1 *target vector cloning*

To clone the *CDX1* cDNA fragment into the target pBI-CMV3 vector, 1 µg pRc/CMV-CDX1 was digested with 2.0 units of the *Not*I restriction enzyme; terminal overhangs were filled in by the Klenow reaction. The linearized plasmid was further digested with

2.0 units of the *Hin*dIII restriction enzyme, gel-purified and quantified to isolate the *CDX1* cDNA fragment. The target vector was prepared by digesting 1 µg pBI-CMV3 with 2.0 units each of *Hin*dIII and *Eco*RV restriction enzymes, followed by purification and quantification. The cDNA fragment insert was ligated into the target vector, transformed into bacteria, extracted and quantified. After sequencing the resulting plasmid, pBI-CMV3-CDX1, using the 5'-TGACGCAAATGGGCGGTAGG-3' forward and 5'-ACCTCTAGAAATGTGGTATGGCTGA-3' reverse primers to confirm identity and orientation of *CDX1* cDNA insert, a plasmid DNA stock was synthesized.

2.5.5.3 BMI1 *target vector cloning*

To clone the *BMI1* cDNA fragment into the target pBI-CMV3 vector, 1 µg pCMV-XL5-BMI1 was digested with 2.0 units each of the *Sma*I and *Sac*I restriction enzymes, purified, blunted with the Klenow fill-in-reaction, gel-purified and quantified to isolate the *BMI1* cDNA fragment. The target vector was prepared by digesting 1 µg pBI-CMV3 with 2.0 units of *Eco*RV restriction enzyme, followed by purification and quantification. The cDNA fragment insert was ligated into the target vector, transformed into bacteria, extracted and quantified. After sequencing the resulting plasmid, pBI-CMV3-BMI1, using the previously detailed sequencing primers to confirm identity and orientation of the *BMI1* cDNA insert, a plasmid DNA stock was synthesized.

2.5.6 Cell cycle analysis

Cellular DNA was stained with the Hoechst 33342 (Invitrogen) vital stain to obtain cell cycle profiles of cell populations. Cells were filtered through a CellTrics® 20-µm filter to obtain single-cell suspensions, divided into aliquots of 10^6 cells, resuspended in 1 mL 70% ethanol and incubated overnight at 4 ºC to fix. The resulting fixed cells were

collected by centrifugation, washed with PBSA, suspended in 1 mL Hoechst staining solution (0.1% Triton X-100 and 2 µg/mL Hoechst 33342 in PBSA) and incubated for 20 min at 37 ºC.

Cell cycle profiles were assessed by flow cytometry on a CyAn ADP analyzer (Dako, Fort Collins, Colorado, USA) using the integrated Summit data collection software. At a forward scatter gain of 1.0 and a side scatter gain of 1.0 and voltage 500 V, doublets and debris were excluded from the analysis using pulse width. At a 405-nm violet laser gain of 1.0 and voltage 500 V, Hoechst 33342 fluorescence was detected; the fluorescence areas were plotted on histograms to identify the G_1, S and G_2/M regions for cell cycle phase analysis.

2.6 Protein analysis

2.6.1 Antibodies

2.6.1.1 Antibodies for Western blot and immunofluorescence

The mouse anti-human monoclonal unconjugated primary antibodies and their experimental dilutions for Western blotting and immunofluorescence included those against BMI1 (Clone 1.T.21, 1:500) (Abcam, Cambridge, UK), β-tubulin (Clone 2-28-33, 1:1000) (Invitrogen), CD55 (Clone IA10, 1:500) (BD Biosciences), CDX1 (1:1,000) (CIL-produced), chromogranin A (Clone LK2H10 + PHE5) (Abcam), EpCAM (Clone AUA1, 1:500) (Cancer Research Technology, London, UK), mucin (Clone PR4D4, 1:500) (Cancer Research Technology) and myofibroblast extract (Clone PR2D3, 1:500) (Cancer Research Technology).

The rabbit anti-mouse polyclonal horseradish peroxidase (HRP)-conjugated secondary antibody used for Western blotting was purchased from Dako and used at a 1:1,000 dilution. The goat anti-mouse polyclonal HRP-conjugated secondary antibody used for immunofluorescence was purchased from Invitrogen and used at a 1:100 dilution.

2.6.1.2 Antibodies for flow cytometry

The mouse anti-human monoclonal conjugated primary antibodies for flow cytometry included those against CD24 (Clone SN3, PE), CD44 (Clone MEM-85, FITC), CD55 (Clone IA10, APC) and EpCAM (Clone B29.1/VU-ID9, FITC) (Abcam). The respective I_GG_{2a} isotype controls include mouse monoclonal antibodies (Clone X40) conjugated to APC, FITC and PE. All antibodies were used at a dilution of 1:10 and were purchased from BD Biosciences if not specified.

2.6.2 Western blot

2.6.2.1 Cell lysate preparation

Cells were detached by scraping, washed with cold PBSA and collected by centrifugation. Cell pellet was resuspended in 1 mL RIPA lysis buffer (150 m*M* sodium chloride, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 50 m*M* tris, pH 8.0 and 1 tablet cOmplete Protease Inhibitor Cocktail, EDTA-free (Roche Applied Science, Burgess Hill, UK)) for 30 min on ice. Remaining cellular debris in lysate supernatant was removed by centrifugation.

2.6.2.2 Lysate protein quantification

The Pierce® BCA Protein Assay kit (Thermo Scientific) was used to quantify lysate protein concentration. Briefly, dilutions of the BSA protein standard were prepared and

transferred in 25-µL triplicate aliquots to a 96-well plate alongside 25-µL triplicate aliquots of cell lysate. After preparing the BCA working reagent (1 mL Reagent B in 50 mL Reagent A) and transferring it in 200-µL aliquots to each well, the plate was incubated for 30 min at 37 ºC. Subsequently, well absorbance was read at 562 nm on a µQuant Microplate Reader and analyzed to construct a standard curve and interpolate the concentrations of the lysates.

2.6.2.3 Polyacrylamide gel electrophoresis

Cell lysates were resolved by polyacrylamide gel electrophoresis. The 10% resolving gel was prepared by combining 3.3 mL 30% acrylamide (Bio-Rad Laboratories), 2.5 mL 1.5 *M* tris pH 8.0, 100 µL 10% SDS, 4 µL TEMED, 100 µL ammonium persulfate (APS) and 4.0 mL water. The resulting solution was poured into a mold to set. The 5% stacking gel was prepared by combining 830 µL 30% acrylamide, 1.75 mL 1.0 *M* tris pH 6.8, 50 µL 10% SDS, 2 µL TEMED, 50 µL APS and 2.83 mL water and overlaying the resulting solution onto the resolving gel to set.

Resulting gel cassette was placed in a Mini Protean II electrophoresis unit (Bio-Rad Laboratories) and overlaid with running buffer (25 m*M* tris, 192 m*M* glycine and 0.1% SDS). Protein lysates were apportioned in 10-µg units, combined at a ratio of 5 to 1 with Laemmli sample buffer (30 mL 1 *M* tris pH 6.8, 50 mL glycerol, 20 mL β-mercaptoethanol, 10 g SDS and 0.25 g bromophenol blue), boiled for 5 min and loaded onto the gel alongside Kaleidoscope pre-stained protein standards ladder (New England Biolabs). Samples were run for 90 min at a constant 120 V.

2.6.2.4 Protein transfer

Polyacrylamide gels were removed from cassettes and soaked in transfer buffer (25 m*M* tris, 192 m*M* glycine and 10% methanol) for 5 min. Transfer cassettes (Bio-Rad Laboratories) were assembled from sponges, filter paper (Whatman, Maidstone, UK), Hybond-C Extra nitrocellulose (GE Healthcare, Little Chalfont, UK) and the gel. Cassettes were placed in a Mini Protean II transfer unit (Bio-Rad Laboratories) and overlaid with transfer buffer. Transfers were run for 60 min at a constant 75 V.

2.6.2.5 Blocking, primary and secondary antibody probing

Nitrocellulose membranes were removed from transfer cassettes, washed in TBST buffer (0.1% Tween-20 in tris buffered saline), blocked for 1 hr in 10 mL blocking buffer (5% dried milk in TBST) and incubated overnight at 4 ºC in 10 mL blocking buffer containing an appropriate dilution of primary antibody. Membranes were washed with TBST 3 times for 5 min each and incubated for 1 hr at room temperature in 10 mL blocking buffer containing an appropriate dilution of secondary antibody. Membranes were washed with TBST a further 3 times for 5 min each.

2.6.2.6 Imaging

Antibody signal was amplified using the ECL Plus Detection kit (GE Healthcare). Briefly, membranes were incubated for 5 min in ECL buffer (25 µL Solution B in 1 mL Solution A) in the dark, imaged using Super RX medical X-ray film (Fujifilm, Tokyo, Japan) and developed on an SRX-201 medical film processor (Konica Minolta, Tokyo, Japan).

2.6.3 Immunofluorescence

2.6.3.1 Cell culture

Cell suspensions were plated on sterilized Number 1 glass cover slips (VWR, West Chester, Pennsylvania, USA) in 6-well plates to yield a final cell count of 500,000 per well. Cells were incubated for 4-7 days to allow cell attachment to cover slips.

2.6.3.2 Fixation and permeabilization

Cells were washed with PBSA, fixed for 10 min at 4 ºC with methanol and for 10 min at 4 ºC with 4% paraformaldehyde. Cells were permeabilized for 10 min at 4 ºC with 0.1% Triton X-100.

2.6.3.3 Blocking, primary and secondary antibody probing

The TSA™ Kit #22 with HRP-Strepdavidin and Alexa Fluor® 488 Tyramide (Invitrogen) was used to probe for protein expression in fixed, permeabilized cells. Briefly, cells were blocked for 1 hr in 250 µL blocking buffer (1% Component D goat serum in PBST) and incubated for 1 hr at room temperature in 100 µL blocking buffer containing an appropriate dilution of primary antibody. Cells were washed with PBST 3 times for 5 min each and incubated for 30 min at room temperature in 100 µL blocking buffer containing a 1:100 dilution of Component C secondary antibody. Cells were washed with PBST a further 3 times for 5 min each.

2.6.3.4 Tyramide amplification, mounting and imaging

Antibody signal was amplified by incubating cells with 100 µL tyramide solution (1:100 dilution of Component A Alexa Fluor® 488 tyramide in a 1:4,000 dilution Component F hydrogen peroxide in Component E amplification buffer) for 5 min at room temperature

in the dark. Cells were washed with PBST 3 times for 5 min and air-dried for 10 min. Cover slips were mounted onto Snowcoat® glass slides (Surgipath, Richmond, Illinois, USA) with 10 µL VECTASHIELD® with DAPI (Vector Laboratories, Burlingame, California, USA), sealed and imaged on an Axiovert 2 fluorescent microscope (Carl Zeiss). Images were collected and analyzed using the integrated Axiovision image processing software.

2.6.4 Flow cytometry

2.6.4.1 Antibody staining and fixation for analysis

Cells were filtered through a CellTrics® 20-µm filter to obtain single-cell suspensions, divided into aliquots of 10^6 cells, washed with wash buffer (1% FBS and 0.1% sodium azide in PBSA) and incubated in 100 µL wash buffer containing an appropriate dilution of primary antibody or isotype control for 30 min at 4 ºC. Cells were washed with wash buffer, re-probed with another antibody if appropriate and washed again with wash buffer. Cells were resuspended in 500 µL 1% paraformaldehyde and incubated overnight at 4 ºC to fix.

2.6.4.2 Flow cytometric analysis

Protein expression profiles were assessed by flow cytometry on a CyAn ADP analyzer using the integrated Summit data collection software. At a forward scatter gain of 1.0 and a side scatter gain of 1.0 and voltage 500 V, doublets and debris were excluded from the analysis using pulse width. At a 488-nm blue laser gain of 1.0 and voltage 500 V, FITC and PE fluorescence was detected; at a 633-nm red laser gain of 1.0 and voltage 750 V, APC was detected. Fluorescence logarithms were plotted to identify negatively- and positively-staining populations.

2.6.4.3 Antibody staining for sorting

Cells were filtered through a CellTrics® 20-µm filter to obtain single-cell suspensions, divided into aliquots of 10^6 cells, washed with serum-free media, collected via centrifugation at 1,200 rpm for 5 min and incubated in 100 µL media containing diluted primary antibody or isotype control for 30 min at 4 ºC. Cells were washed with media, re-probed with another antibody if appropriate, washed again with media and collected via centrifugation at 1,200 rpm for 5 min. Cells were resuspended in 500 µL media and stained supravitally with 20 µL propidium iodide (PI) (50 µg/mL, BD Biosciences).

2.6.4.4 Fluorescence-activated cell sorting (FACS)

Cells were sorted by flow cytometry on a MoFlo™ cell sorter (Dako) using the Summit data collection software. At forward and side scatter gains of 1.0 and voltage 500 V, doublets and debris were excluded using pulse width. At a 488-nm blue laser gain of 1.0 and voltage 500 V, PI fluorescence was used to exclude dead cells. The 488-nm laser was further used to detect FITC and PE fluorescence; at a 633-nm red laser gain of 1.0 and voltage 750 V, APC was detected. Fluorescence logarithms were plotted to identify negatively- and positively-staining populations.

The top and bottom 1% of the fluorescence distributions were sorted and collected in 1 mL of serum-free media for subsequent experimentation. For two-color experiments, wedge-shaped gates corresponding to 1% of the distributions were constructed to sort the extremes of the fluorescence distributions. For three-color experiments, the top and bottom 10% of the APC fluorescent distribution were sorted; wedge-shaped gates corresponding to the top and bottom 10% of the resulting two APC distributions (Figure 2.1) were constructed to sort the 1% extremes of the entire distributions.

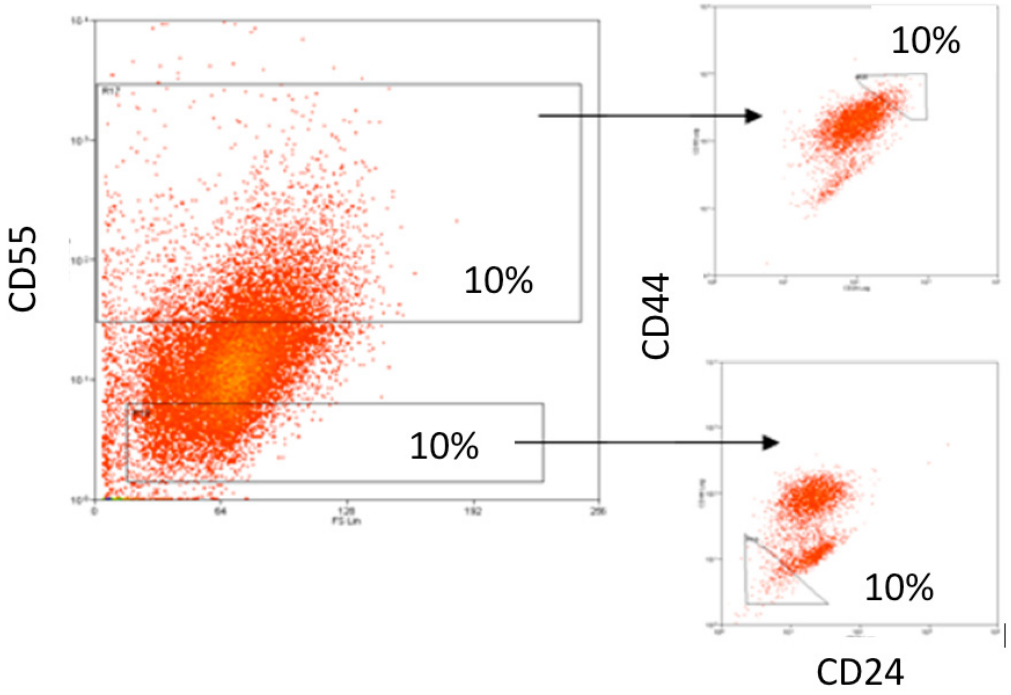


Figure 2.1 Flow cytometry dotplots illustrating gating logic of a three-color cell sort for the C80 cell line. The sort illustrated here isolated the top and bottom 1% of the three-color distribution (corresponding to the $CD24^{high}CD44^{high}CD55^{high}$ and $CD24^{low}CD44^{low}CD55^{low}$ populations) by selecting the top 10% wedge of the CD24/CD44 plot generated from the top 10% slab of the CD55 plot and by selecting the bottom 10% wedge of the CD24/CD44 plot generated from the bottom 10% slab of the CD55 plot.

2.7 Plasmid transfection

Cell suspensions were plated in 24-well plates to yield a final cell count of 10^5 per well and incubated for 24 hrs. The Lipofectamine™ LTX with Plus™ reagent kit (Invitrogen) was used to transfect plasmid DNA into cells. Briefly, 1 µg plasmid DNA was diluted in 100 µL Opti-MEM Reduced Serum medium, combined with 1 µL PLUS™ reagent and incubated for 15 min at room temperature. Opti-MEM Reduced Serum medium is a serum-free medium containing a proprietary mixture of salts that is thought to promote greater transfection efficiency; PLUS™ reagent is another proprietary mixture of salts also thought to promote greater transfection efficiency. The resulting solution was

combined with 5 µL Lipofectamine™ LTX reagent and incubated for 30 min at room temperature. Subsequently, the solution was added drop-wise to one well; the cells were incubated overnight to allow expression of the transfected plasmid.

Transfected cells were imaged on an Axiovert 135 inverted fluorescent microscope (Carl Zeiss). Images were collected and analyzed using the integrated Axiovision image processing software.

2.8 Data analyses

2.8.1 General data analyses

Data analyses, including mean and standard error determination as well as curve fitting, were conducted using the GraphPad Prism® software package (GraphPad Software, La Jolla, California, USA).

2.8.2 Microarray expression analyses

Total RNA from thirty colorectal cancer cell lines, namely C80, C84, C99, CC20, CCK81, Colo741, DLD1, Gp2D, HCA7, HCT116, HRA19, HT29, LIM1863, LS174T, LS180, LS411, LS413, Lovo, NCIH716, RCM1, RKO, SKCO1, SNUC2B, SW1222, SW1417, SW48, Vaco5 and WIDR, was extracted, reverse-transcribed to cDNA and hybridized to Human Genome U133 Plus 2.0 (Affymetrix, Santa Clara, California, USA) microarray chips by members of the CIL. Resulting gene expression data were grouped according to exogenous characteristics, such as three-dimensional colony morphology; expression means of groups were compared to each other by Student's *t*-tests using the Partek Genomic Suite software package (Partek, St. Louis, Missouri, USA). The

Bonferroni correction was applied to mean comparisons to reduce the number of false significant results. A fold-change of 1.5 or greater at *p*-value of less than or equal to 0.02 was considered statistically significant.

2.8.3 Microarray expression classification algorithm

To classify cell lines on the basis of differential gene expression, a statistical algorithm, developed in the CIL, was applied to the expression data. Briefly, cell lines were organized in ascending order based on their respective expression levels for one gene transcript and assigned rankits, expected values from the normal distribution based upon rank order. A linear regression of rankit versus rank was constructed and compared to the true plot of gene expression versus rank; statistically significant deviations from the regression are indicative of a new gene expression classification, distinct in its expression level from classifications prior.

CHAPTER 3

MICROARRAY EXPRESSION COMPARISONS OF COLONY MORPHOLOGIES

3.1 Introduction

The differentiation status of a particular colorectal tumor has been a well-studied subject. Under the microscope, pathologists often characterize a tumor by the degree of organization of the cells, with tumors showing intercellular structure and polarity deemed "well-differentiated" and tumors showing little organization "poorly-differentiated." Studies assessing the usefulness of differentiation status in tumor biology and clinical outcome have been numerous, and such diverse features such as peritoneal dissemination (Yang, et al., 2004), apoptosis-related biomarker expression (Kontos, et al., 2008) and presence of the putative cancer stem cell markers ABCA5 and ABCB1 (Ohtsuki, et al., 2007) have been linked to differentiation status.

Interestingly, cell lines have also been shown to generate colonies with variable differentiation status when grown in three-dimensional matrices. A large-scale study by Kenny and colleagues (2007) focusing on breast cancer cell lines identified four major classes of colony morphology characterized by differences in cell-to-cell contacts and cellular polarity. Furthermore, these morphology classes were linked to expression of growth factor receptors and membrane receptors important for adhesion and invasion. Within colorectal cancer, the SW1222 cell line can differentiation *in vitro*, and such differentiation is controlled by an extracellular matrix membrane receptor (Pignatelli and Bodmer, 1988). There is a relationship between growth behavior and biochemical characteristics, and such characteristics have relevance to tumorigenesis. An assessment of growth behavior could prove useful in identifying novel molecules that are not only linked to colony morphology but also other factors, such as cancer stemness or differentiation.

3.2 Results

3.2.1 Colony morphology assay

To assess the diversity of colony morphologies in a panel of colorectal cancer cell lines, thirty cell lines were selected based on the availability of gene expression microarray data—such data made comparisons between the cell lines on the basis of biochemical makeup possible. The cell lines, viz. C80, C84, C99, CC20, CCK81, Colo741, DLD1, Gp2D, HCA7, HCT116, HCT15, HRA19, HT29, LIM1863, Lovo, LS174T, LS180, LS411, LS513, NCIH716, PCJW, RCM1, RKO, SKCO1, SNUC2B, SW1222, SW1417, SW48, Vaco5 and WIDR, were originally selected because they demonstrate the diversity of colorectal cancer cell lines, thus recapitulating the heterogeneity of human tumors.

To assess growth characteristics within a three-dimensional matrix, all thirty cell lines were grown in a mixture of Matrigel™ and collagen I. In the past, three-dimensional growth has been assessed in soft agar, but the polysaccharide makeup of the matrix does not reflect the true components that cells would experience within the extracellular matrix (ECM). At the same time, Matrigel™ on its own is quite soft and gelatinous, making handling difficult. By mixing collagen I with Matrigel™, the resulting stiffer solution still possesses the ECM proteins that could enhance colony formation.

After incubating the cells in serum-containing media for 2 weeks, colonies were imaged to assess the size, shape and organization of any grown colonies. Each cell line gave rise to unique colonies, but common morphologies emerged (Figure 3.1). Four distinctive structures emerged—the network, dense, lumen and intermediate—each characterized by intracolony organization, colony border definition and colony density. Cell lines were grown in a pure Matrigel™ matrix to confirm the presence of the observed morphologies.

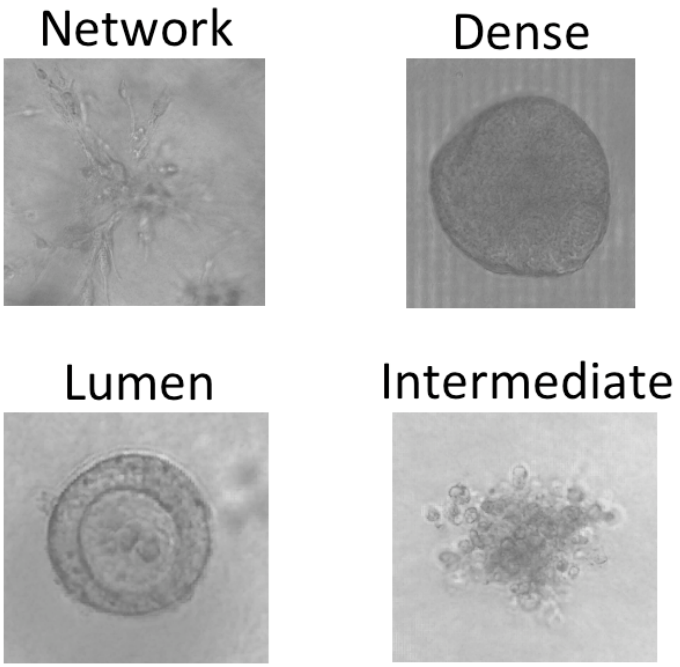


Figure 3.1 Photographs of three-dimensional growth cultures of cell lines illustrating the fibrous, nonspherical network morphology (Colo741), dense, spherical dense morphology (HCT15), organized, polarized lumen morphology (LS513) and diffuse, undefined intermediate morphology (Lovo). All photographs were taken at 10x under phase contrast.

A minority of cell lines gave rise to network colonies, characterized by elongated, fibrous structures with small nodes that formed a network of cells. There were few cell-to-cell contacts, and agglomerations of cells did not radiate spherically away from the clonal cell of origin. Cell polarity appeared to be nonexistent.

A larger number of cell lines gave rise to dense colonies, characterized by well-defined colony borders and densely-packed interiors. Cells within the colony appeared undifferentiated and similar in morphology to each other, as though the colony was a packed spheroid of clonal cells. Little cell polarity was observed. Sizes varied, but dense colonies tended to be largest of all other colonies, implying higher levels of proliferation.

An even larger number of cell lines gave rise to lumen colonies. As opposed to the other colony morphologies, the lumen morphology was characterized by the presence of a large

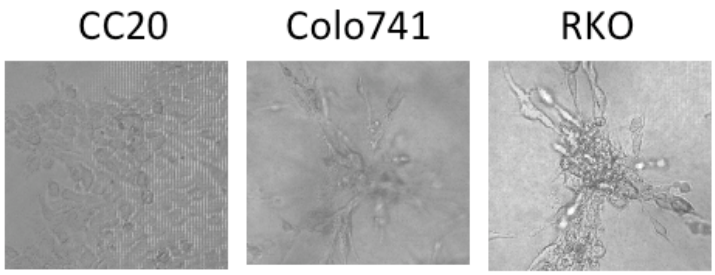
central lumen within the colony. The borders of the lumen and of the colony are well defined, and cells appeared to be polarized around the lumen upon light microscopic inspection. Individual cells could be identified, and there is more variability in individual cell morphology, as opposed to the lack of variability in dense colonies. The organization of the colonies was striking in that it resembles that of the normal colonic crypt—layers of polarized epithelial cells surrounding a luminal space.

The final, most common, colony was of an intermediate morphology. Colony borders were somewhat defined and possessed stellate cell protrusions. Cell density within the colony was not high, and some cell polarity and heterogeneity in shape was observed. Intermediate colonies tended to be quite variable in size. Intermediate colonies could be characterized as a hybrid of the spheroid, defined dense colonies and the stellate, fibrous network colonies.

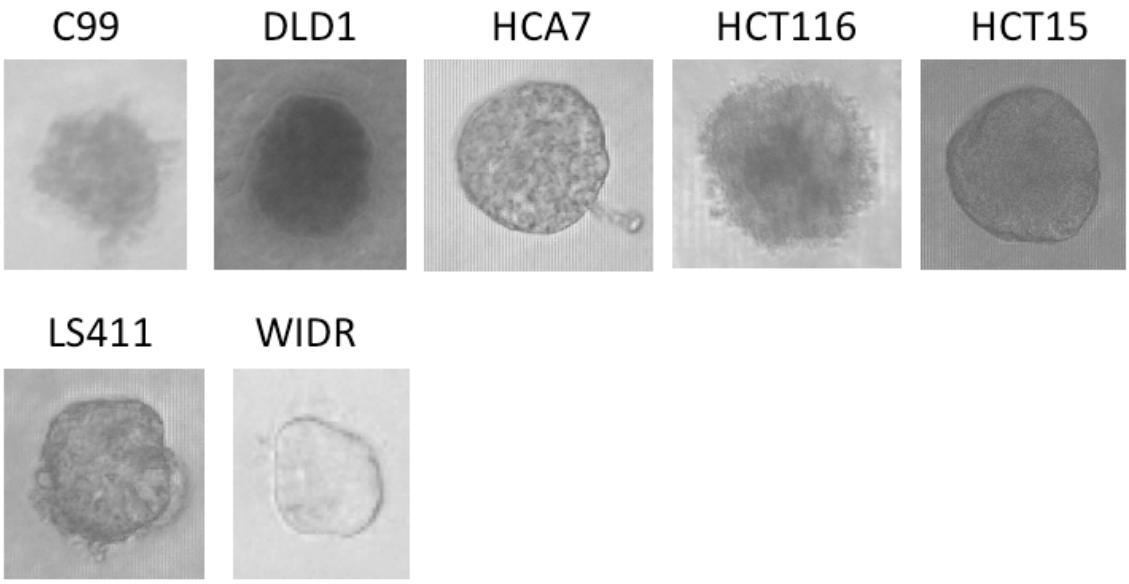
3.2.2 Distribution of colony morphologies

Cell lines were grouped according to the proportions of extant colony organizations (Figure 3.2). In most cases, a single morphology dominated a cell line and accounted for over 75% of the total colonies.

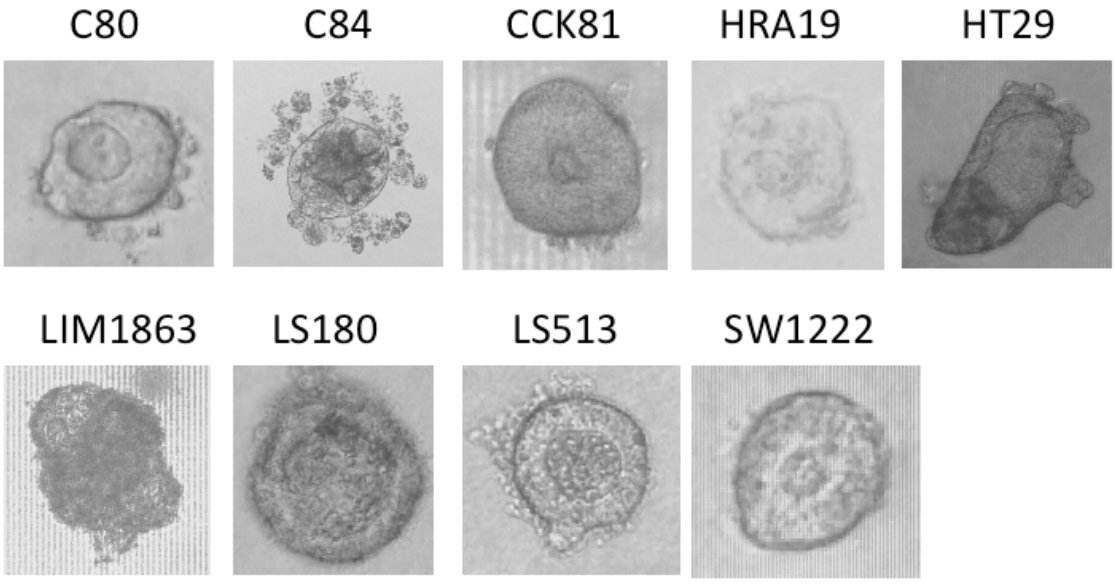
(a)


CC20
Colo741
RKO

(b)


C99
DLD1
HCA7
HCT116
HCT15
LS411
WIDR

(c)


C80
C84
CCK81
HRA19
HT29
LIM1863
LS180
LS513
SW1222

(d)

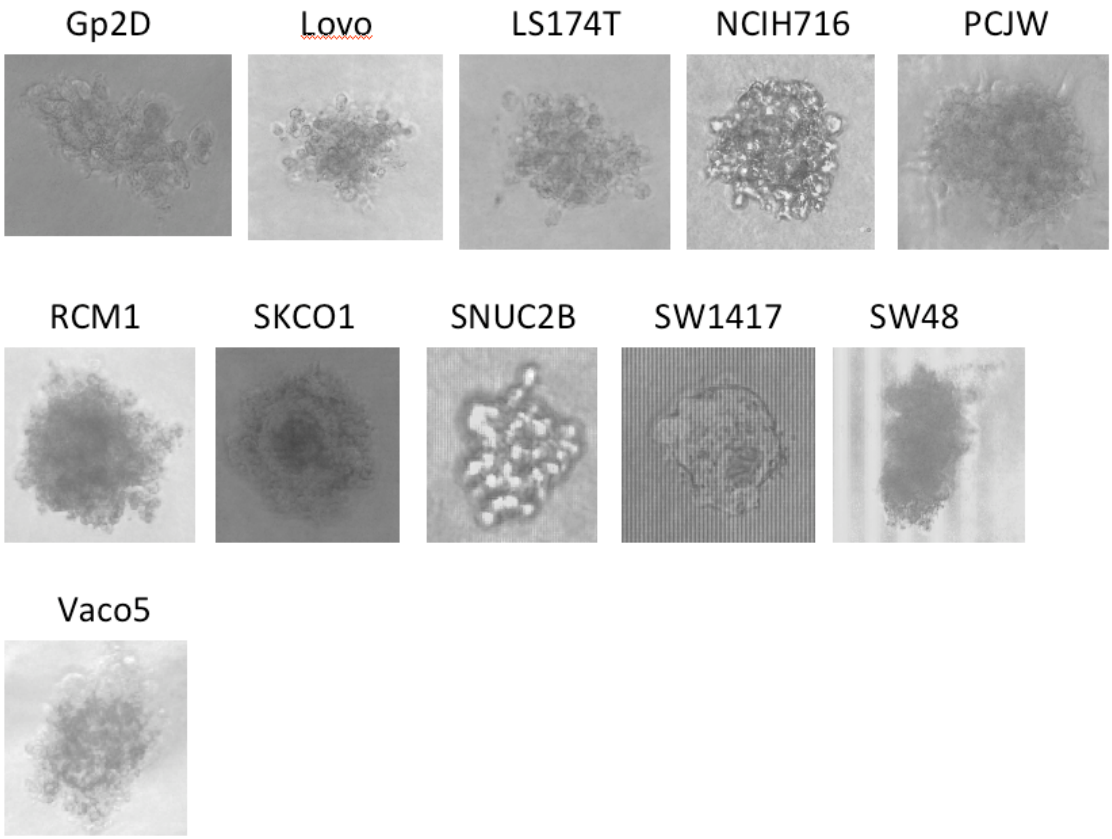


Figure 3.2 Photographs of three-dimensional growth cultures of cell lines sorted into the four major colony morphologies, namely the network (a), dense (b), lumen (c) and intermediate (d). All photographs were taken at 10x under phase contrast.

The three cell lines that produce network colonies, namely CC20, Colo741 and RKO, look remarkably similar in terms of their colony appearance—all lines produced very similar stellate, branching agglomerations of cells with little spherical organization.

The seven cell lines that produced dense colonies, namely C99, DLD1, HCA7, HCT116, HCT15, LS411 and WIDR, also look remarkably similar in terms of their colony appearance—all the lines produced spherical, dense colonies with defined borders. There were variations in size among the cell lines, with DLD1, HCT116 and HCT15 producing large, numerous colonies and C99 and HCA7 producing small and rare colonies. LS411 and WIDR colonies were of moderate size and frequency.

The most interesting differences in morphology were found among the nine cell lines that gave rise to lumen colonies, namely C80, C84, CCK81, HT29, HRA19, LIM1863, LS180, LS513 and SW1222. While colonies within each cell line generally appeared to correspond to the model of a colony containing polarized cells surrounding a central lumen, the sizes and shapes of the colonies varied considerably. A high proportion of colonies containing lumens was observed in the C80 cell line, while C84, CCK81 and LIM1863 produced low numbers of lumen-containing colonies. The C80, HRA19, LS513 and SW1222 cell lines produced spherical colonies with spherical lumens, while HT29 and LS180 produced lobular colonies with irregularly shaped lumens. C80, HRA19 and LIM1863 produced large colonies, while C84 and LS513 produced small colonies. The HRA19 cell line was distinctive in that the central lumens were large relative to the colony size.

Finally, the eleven cell lines that produced intermediate colonies resemble each other when grown in three dimensions. There were some differences—a potential dichotomy within the intermediate population could be devised based on the cell density of the colonies. Some cell lines, including SKCO1, SW1417 and SW48, are denser than others, including Gp2D, Lovo and SNUC2B. Other than density, proportions of colonies with stellate protrusions and poorly defined borders were similar across all eleven lines.

3.2.3 Gene expression microarray analysis

3.2.3.1 Colony morphology and differentiation status

Of the classifications, the two with the most striking differences were the dense and lumen classifications. Save for the cell density and border definition, dense colonies were unpolarized and undifferentiated, while lumen colonies were polarized and differentiated.

These points of difference may imply a link between the morphology of the colony and differentiation status. Dense colonies appeared to be representative of cells that remain in an undifferentiated state and do not respond to differentiation signals from the surrounding growth matrix or medium. Lumen colonies, conversely, appeared to be representative of cells that are still able to respond to differentiation signals to give rise to organized structures reminiscent of the normal human colonic crypt. In summary, dense colonies can be thought of as pools of undifferentiated cells, while lumen colonies can be thought of as pools of differentiated cells. Indeed, recent experiments have found lumen colonies to contain cells representing the three major lineages of colon—CDX1-, chromogranin A- and mucin-expressing cells were identified, corresponding to the columnar, enteroendocrine and goblet cell lineages (Yeung, et al., 2010).

3.2.3.2 Microarray overview

To understand the regulation of stemness and differentiation in colorectal cancer, comparing the dense and lumen cell line groups on the basis of gene expression may prove useful in understanding which genes are differentially expressed. Therefore, microarray data previously collected in the CIL was partitioned according to cell line colony morphology, and multiple Student's *t*-tests were performed for each gene transcript probe comparing mean expression for the dense lines versus that of the lumen lines. To ensure that identified gene transcripts actually represent those that are differentially expressed, the Bonferroni correction was applied to the statistical tests in order to exclude false positives. To include only meaningful gene differences, genes were investigated if the (corrected) *p*-value was less than 0.02 and the fold-change between the two morphology groups was 1.5 or larger.

The resulting statistical analysis identified 822 gene transcripts to be differentially expressed, with 394 probesets (Appendix B, Table B.1) showing significant upregulation in dense cell lines and 428 probesets (Appendix B, Table B.2) showing upregulation in lumen cell lines. A more detailed inspection of the 822 genes revealed the impact of genes known to be influential in colorectal development. At the same time, multiple genes found to be differentially expressed had few, if any, known functions in the colon. A careful examination of the extant information on these genes would prove useful in understanding which genes to investigate further.

3.2.3.3 Genes upregulated in dense lines

Of the genes that tended to be expressed more strongly in the dense lines, seven stand out because of their roles in carcinogenesis, tissue development and stem cell biology: *RUNX2*, *BMP1*, *FGF2*, *ITGB8*, *ALDH1A3*, *BMI1* and *CD55*.

Overexpressed 6.8-fold at a *p*-value of 0.0002, *RUNX2* is one of the genes most strongly associated with the formation of dense colonies. The associated protein product, Runx2, is a transcription factor known to be critical in bone development by promoting the maturation and differentiation of cartilage-forming chondrocytes and bone-forming osteoblasts (Pratap, et al., 2006) through interactions with other transcription factors and cell signaling proteins. In breast and prostate cancers, Runx2 is associated with higher metastatic potential and oncogenesis (Pratap, et al., 2006)—in murine colorectal cancers, Runx2 stimulates the transcription of the metastatic osteopontin gene (*OPN*) (Wai, et al., 2006). Although it promotes differentiation in the bone, its effects in solid tumors are much more pro-proliferative and pro-invasion.

Another gene target of Runx2 activity is *BMP1*, which is itself overexpressed 1.6-fold at a *p*-value of 0.0013. Its extracellular metalloproteinase protein product, bone morphogenetic protein 1 (BMP1), also mediates bone development by activating procollagens I-III via C-propeptide cleavage to promote the formation of the ECM (Hopkins, et al., 2007). However, in contrast to the effects of Runx2, Wu and colleagues (2008) have found that BMP1 is induced upon pharmaceutical inhibition of HCT116 colorectal cancer cell line clonogenicity. In contrast to other BMPs, many of which drive increased tumorigenicity and invasiveness (particularly BMP2 and BMP7) (Thawani, et al., 2010), the evidence for BMP1 seems to indicate that BMP1 expression is associated with a dampening of cancer proliferation and activity. Thus, its association with proliferative, undifferentiated dense colonies is puzzling.

The remaining genes all have potential roles related to the cellular promotion of a stem-like phenotype. *FGF2*, overexpressed 2.5-fold at a *p*-value of 0.0044, encodes the basic fibroblast growth factor (bFGF), a secreted glycoprotein that readily binds to the ECM and to fibroblast growth factor receptors, including FGFR2, to drive downstream signaling events via the MAPK pathway (Turner and Grose, 2010). In normal tissue, including the colon, bFGF is important for wound healing and neovascularization, but these pro-proliferative effects become dysregulated in many solid tumors to promote cancer growth and invasiveness (Turner and Grose, 2010). Within the context of development, bFGF promotes a stem cell phenotype and is necessary for maintenance of pluripotent stem cell cultures in vitro (Navarro-Alvarez, et al., 2008); indeed, as previously discussed, bFGF is used to maintain *in vitro* cultures of cancer stem cells.

ITGB8, overexpressed 1.5-fold at a p-value of 0.0046, is a member of the integrin β-chain gene family and encodes the β8 integrin protein. Although little is known about the

specific functions of the transmembrane glycoprotein β8 integrin apart from a study linking its expression to an increased propensity for breast tumors to metastasize to the lung (Culhane and Quackenbush, 2009), other integrin proteins function as sensors of extracellular activity with respect to cell-to-cell adhesion, basement membrane attachment, cell motility and ECM interaction (Rathinam and Alahari, 2010), especially with respect to interactions with collagen (Pignatelli and Bodmer, 1989). Thus, multiple integrin family members, including α3, α6, β1 and β4 can drive tumor growth and invasiveness when dysregulated (Rathinam and Alahari, 2010). Furthermore, expression of the α2β1 integrin complex has been found to be associated with the cancer stem cell phenotype in the HRA19 colorectal cancer cell line (Kirkland and Ying, 2008). However, work by Pignatelli and colleagues (1990) links the loss of the α2, α3 and β1 integrin chains of the collagen receptor to a loss of differentiation in primary colorectal tumor tissues, so the true relationship between integrin expression and cancer differentiation has yet to be fully elucidated.

Similarly, *ALDH1A3*, which was overexpressed 4.5-fold at a *p*-value of 0.0080, has not been extensively studied, but the related gene *ALDH1A1* and its associated cytosolic protein product have been found to mediate the synthesis of retinoic acid, a small molecule regulator of stemness (Su, et al., 2010). Unsurprisingly, *ALDH1A1* has been associated with the cancer stem cell population in breast (Ginestier, et al., 2007) and bladder cancer (Su, et al., 2010), among other solid tumors, so it seems logical that gene family member *ALDH1A3* is upregulated in the undifferentiated dense colony cell lines.

The final two genes, *BMI1* (1.5-fold overexpression at a *p*-value of 0.0019) and *CD55* (2.8-fold overexpression at a *p*-value of 0.0155), as previously discussed, have well-documented links to the maintenance of the stem cell phenotype and correlations with the

cancer stem cell population. Because the biology of *BMI1* and *CD55* are well established, assessing the roles of these two genes would likely be fruitful in understanding more fully the control mechanisms of cancer stem cells in colorectal cancer.

3.2.3.4 Genes upregulated in lumen lines

In contrast to the genes more strongly expressed in dense lines, many genes upregulated in lumen lines have been extensively studied with respect to their functions in the colon and in cancer. Among many, seven are highlighted for their specific links not only to colonic crypt regulation but also to colorectal carcinogenesis: *FABP1*, *MUC3A*, *CEACAM5*, *GPA33*, *TFF3*, *VIL1* and *CDX1*.

The most strongly overexpressed gene in lumen lines was *FABP1* (17.6-fold overexpression at a *p*-value of 0.0027), a gene that encodes the cytosolic liver fatty acid binding protein (L-FABP). Operating physiologically to bind long-chain fatty acids within cells to ensure their proper transport and metabolism (Besnard, et al., 2002) and structurally similar to a related protein expressed in the small intestine, intestinal FABP (I-FABP), L-FABP has been associated with a downregulation of tumorigenicity. A study of the adenoma-carcinoma colorectal cancer sequence highlighted the downregulation of expression of FABP1 as normal colon tissue progresses toward a carcinoma (Lee, et al., 2006). Immunohistochemical studies of colorectal tissue, both normal and cancerous, confirms the expression findings—carcinomas expressed very little L-FABP versus normal tissue, and such a downregulation of L-FABP expression was correlated with poorly-differentiated tumors (Lawrie, et al., 2004). Thus, strong circumstantial evidence exists for a functional role of *FABP1* in the differentiation status of colorectal cancer, a role implied by the strong correlation of *FABP1* to the formation of lumen colonies.

Another gene with a relationship to colorectal physiology is *MUC3A*, overexpressed 2.7-fold at a *p*-value of 0.0024. The associated secreted protein, mucin 3A, like other mucins, represents part of the glycoprotein constituent of mucus and helps protect epithelial surfaces from chemical insults (Byrd and Bresalier, 2004)—colonic epithelial cells in particular secrete mucin 3A. Within the context of colorectal cancer, little is known regarding its role. Studies of mRNA and protein expression in tumors have found a downregulation of *MUC3A* mRNA as compared to normal tissue but a paradoxical upregulation in mucin 3A protein staining. Since mucin 3S is heavily glycosylated, it has been suggested that this increase in protein staining is attributable to deglycosylation events in the tumor that reveal more antibody binding sites (Ho, et al., 1993). Thus, *MUC3A* appears to be negatively associated with carcinogenesis, implying a possible role for the gene and protein in dampening the progression toward a tumor.

Perhaps the most well-studied protein within the context of colorectal cancer is carcinoembryonic antigen (CEA), the transmembrane protein encoded by the *CEACAM5* gene, overexpressed 7.8-fold at a *p*-value of 0.0083. Although some evidence suggests that it can serve as an intercellular adhesion molecule that mediates cell-to-cell contacts (Kuespert, et al., 2006), it is more widely known as a biomarker of colorectal cancer. CEA is widely overexpressed in colorectal tumors versus normal tissue, and Chan and Stanners (2007) have found its expression has been correlated with a poorer clinical outcome. Paradoxically, induced CEA expression in colorectal cancer cells dampens *in vitro* differentiation and disrupts cell organization (Ilantzis, et al., 2002). Furthermore, CEA expression has been found in cultured colorectal CSCs (Fang, et al., 2010). Consequently, its relationship to differentiated lumen colonies is unknown.

The fourth gene, *GPA33*, was upregulated 6.2-fold at a *p*-value of 0.0095. The surface glycoprotein product, A33, is primarily a marker of colorectal cancer, although Health and colleagues (1997) concluded that A33 is an intestine-specific cell adhesion molecule based on protein structural analysis and immunohistochemical staining of the normal human colonic epithelium. *GPA33* is a target of the PPARγ transcription factor via an interaction with another transcription factor, KLF4 (Rageul, et al., 2009). In colorectal cancer, A33 is expressed on more than 95% of tumors and has been targeted by monoclonal antibody therapy to reduce tumor burden (Chong, et al., 2005), but its presence in two cyclin-dependent kinase 4/human telomerase-immortalized primary colonic epithelial cultures was associated with growth arrest, indicative of differentiated cells (Roig, et al., 2010). While all cell lines assessed on the basis of colony morphology are cancerous, and thus are likely to overexpress A33, it is likely that A33 expression is higher in lumen cell lines because of a greater preponderance of differentiated cells and its association with the CDX1 differentiation marker (Chan, et al., 2009).

Intestinal trefoil factor-3 (TFF3), the protein product of the gene *TFF3* (overexpressed 5.9-fold at a *p*-value of 0.0078) is secreted by the goblet cells of the colon into the lumen. Normally functioning to promote migration of epithelial cells and dampen apoptosis, TFF3 is also expressed in colonic tumors (Babyatsky, et al., 2009), but mRNA expression of TFF3 falls as tumors progress from an early stage to a late stage, implying an inverse relationship between TFF3 and aggressiveness of tumor (Uchino, et al., 1997). Furthermore, an inverse relationship is observed between TFF3 expression and the differentiation status of the tumor, implying an additional role for TFF3 in carcinoma differentiation (Uchino, et al., 1997). However, in the rat colon cancer cell line LPCRI, siRNA knockdown of *TFF3* mRNA expression dampens cell invasion, and human

colorectal cancer liver metastases express TFF3 as well (Babyatsky, et al., 2009). The evidence, therefore, is contradictory regarding the role of *TFF3* in colon cancer—its link to lumen colonies may lend credence to its possible function as a mark of differentiation.

The sixth gene overexpressed in lumen cell lines versus dense lines (4.3-fold overexpression at a *p*-value of 0.0019) was *VIL1*. The associated cytosolic protein product, villin 1, forms part of the cytoskeleton within colonic epithelial brush borders through its calcium-dependent actin modification functions (Xi, et al., 2006). Its functions in colorectal cancer are more vague—a comparative proteomic study of the adenoma-carcinoma sequence identified villin 1 as a protein whose expression is positively proportional to the emergence of carcinomas, although the associated *p*-value with increase in villin 1 was a high 0.05 (Roth, et al., 2010). However, *VIL1* expression is controlled by the colonic differentiation transcription factors CDX1 and CDX2; induction of CDX1 and CDX1 protein expression in the SW480 cell line increased expression of villin 1, while shRNA knockdown of CDX1 and CDX2 dampened villin 1 expression (Yamamichi, et al., 2009). Since the link between the caudal-type homeobox genes and colonic epithelial differentiation is established, it is likely that *VIL1* is also correlated with differentiation, as its association with lumen cell lines would imply.

The final gene, *CDX1* (4.3-fold overexpression at a *p*-value of 0.0006), as previously discussed, has strong empirical evidence linking it to the differentiation of colonic crypt epithelial cells. Since the role of *CDX1* in normal tissue has been studied, assessing its role within the context of cancer stemness and differentiation within lumen and dense colonies would be prudent.

3.3 Discussion

Studies of cancer stem cells are often difficult because of a lack of *in vitro* methodologies for assessing the stem-like or differentiated phenotypes. It has been demonstrated here that the three-dimensional culture of colorectal cancer cell lines can give rise to colonies that possess characteristics of differentiation, including the formation of organotypic structures, polarized cells and multiple colonic epithelial cell lineages, and of stemness, including the formation of dense cell-to-cell structures with little internal structure. More extensive experiments can be performed to confirm and understand more fully the nature of the cell polarization observed within the lumen cell lines, such as immunohistochemical staining of the colonies for the presence of brush borders and microvilli on the apical (internal colony border) side and of integrins on the basolateral (external colony border). Electron microscopy of the colonies themselves can also reveal fine microstructural features characteristic of apical and basolateral membranes, such as the formation of organized E-cadherin junctions with neighboring cells, while additional immunohistochemistry assessments can confirm the presence of concomitant basolateral expression of β-catenin associated with proper E-cadherin functioning (Jou, et al., 1995).

Since it is likely that the expression patterns intrinsic to the two colony morphologies are the major cause of the differences in colony organization, comparisons of mean mRNA expression levels for each transcribed gene would prove useful in understanding which genes are upregulated or downregulated in response to differentiation or maintenance of stemness. Comparing the expression levels of cell lines forming dense colonies and lumen colonies identified 428 potential genes associated with differentiation and 394 genes associated with stemness.

Of these 822 genes, fourteen were highlighted as potential genes of interest for further study. Of the genes correlated with the dense colony phenotype, four have not previously been found to enhance stemness, including *RUNX2*, *BMP1*, *ITGB8* and *ALDH1A3*, while the remaining three have, including *FGF2*, *BMI1* and *CD55*. Since the evidence linking *BMI1* to cancer stemness is strong, probing its effects in a panel of colorectal cancer cell lines would likely produce interesting results. Furthermore, since *CD55* is a cell surface marker, dividing cells into populations of varying expression for *CD55* via flow cytometry would be possible, allowing for in-depth clonogenicity and proliferation assessments that are classically performed to identify cancer stem cells.

Of the genes correlated with the lumen colony phenotype, three are genes not previously associated with colon cell differentiation, including *FABP1*, *MUC3A* and *VIL1*, while the remaining four have substantial experimental data identifying functions within the normal colon as well as in colorectal cancer, including *CEACAM5*, *GPA33*, *TFF3* and *CDX1*. Since mRNA and immunohistochemical data link *CDX1* expression to differentiation of colonic epithelial cells, *CDX1* is a prime candidate for understanding the roles of normal differentiation patterns not only within colorectal cancer but also within cancer stem cells.

Thus, the three genes that will be studied further are *BMI1*, *CD55* and *CDX1*, as well as their associated protein products.

CHAPTER 4

IDENTIFICATION OF CD55 AS A NOVEL MARKER OF CANCER STEMNESS

4.1 Introduction

The morphology microarray expression assessment revealed the gene *CD55* to be potentially linked to the cancer stem cell phenotype by virtue of its high relative mRNA expression levels in undifferentiated dense colony cell lines. The associated transmembrane protein product CD55 (also known as decay accelerating factor, or DAF) was originally identified in the hematological system on erythrocytes (Hoffman, 1969) as an inhibitor of the membrane attack complex of the classical complement pathway. Thus, cells expressing CD55 are able to resist cell cytotoxicity via complement activation (Mikesch, et al., 2006). Within the normal large intestine, CD55 is expressed sporadically—positive cells were generally found on the luminal surface of colonic crypts (Koretz, et al., 1992), but its function in colorectal physiology has not been fully elucidated. Within colorectal tumors, though, CD55 expression was found to be much higher—a significant minority of tumors expressed CD55 ubiquitously, both adenomas and carcinomas (Koretz, et al., 1992).

Because of this heterogeneity of expression, it appears that CD55 is not a universal marker of colorectal cancer cells, so it may have other functions. Evidence from other solid tumors indicates that one of these other functions could be an association with the cancer stem cell (CSC) phenotype. CD55 expression was correlated with the side population phenotype in MCF7 breast cancer cells, and $CD55^{high}$ cells were more resistant to induced apoptosis (Xu, et al., 2007). Knockdown of CD55 expression in PC-3 prostate cancer cells with an siRNA plasmid construct dampened tumor growth in a mouse xenograft model (Loberg, et al., 2006). Chemoresistant ovarian cancer cells expressing high levels of P-glycoprotein, thought to be a CSC marker, have been found to express high levels of CD55 (Odening, et al., 2009). While its specific role in colorectal cancer

may be limited to its effects in inhibiting complement-activated cell lysis or potentially its interaction with the bacterial gut flora, CD55 appears to be a promising protein that could be associated with cancer stemness. Its status as a membrane protein allows for facile flow cytometric separations of cells on the basis of CD55 expression.

Many methods are available for testing the presence or absence of the stem cell phenotype. Since it recapitulates the *in vivo* environment by mimicking the distributions of extracellular proteins, fibroblasts, immune cells and other components of the tumor niche, tumor initiation in a mouse model is the gold standard test of cancer stemness. At the same time, isolating the cells in an *in vitro*, defined system allows for detailed analyses of tumor initiation and growth. By growing isolated cells in three-dimensional growth matrices, the abilities of cells to initiate colonies without uncontrolled external growth cues can be assessed, abilities that perhaps are more indicative of the CSC phenotype than xenografts in a permissive milieu, many of which lack immune cells crucial for a true representation of the colonic crypt. Furthermore, growing isolated cells in plates to assess proliferation allows for a direct observation of the intrinsic abilities of such cells to proliferate without growth cues. Thus, *in vitro* assessments can identify CSCs phenotypically while allowing for more externally controlled measurements.

4.2 Results

4.2.1 *CD55* mRNA expression levels vary among colorectal cancer cell lines

The results of the Bonferroni-corrected Student's *t*-test comparing mean *CD55* mRNA expression levels of dense cell lines to lumen cell lines indicated a statistically-significant increase in expression associated with the formation of dense colonies. To determine if *CD55* expression can actually be partitioned into two groups of cell lines, *CD55* mRNA expression was plotted against cell lines in rank order (Figure 4.1) to assess any discontinuities in the distribution.

Applying the rankit regression algorithm developed by the CIL to the *CD55* expression plot allowed for the separation of the cell line panel. At a significance level of 0.001, one discontinuity in the line plot was observed between the LS513 and Gp2D cell lines at an expression level of 841 fluorescence units (midway between the expression level of LS513, 672, and of Gp2D, 1,009). The nine lumen cell lines tended to appear on the lower end of the distribution, while the seven dense cell lines tended to appear on the higher end.

To test the association between *CD55* expression level and cell line colony morphology, a Fisher's exact test was performed on the partition of dense and lumen cell lines into high and low *CD55* expression classifications (Figure 4.2). Eight lumen lines and two dense lines were low for *CD55*, while five dense lines and one lumen line were high for *CD55*; HT29 was the one lumen line exhibiting high *CD55* mRNA expression in the microarray. The resulting 2 x 2 table indicated that *CD55* expression was positively associated with the formation of dense lines at a *p*-value of 0.0245.

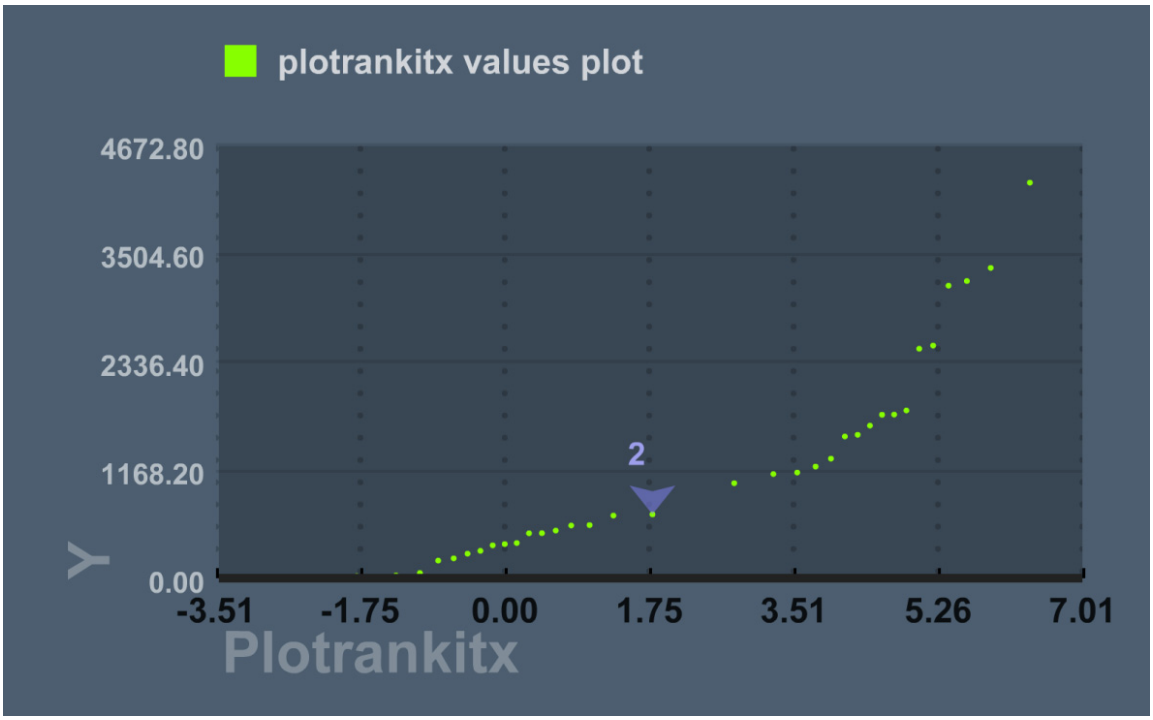


Figure 4.1 Plot of *CD55* microarray expression data versus rankits, highlighting the data range (three orders of magnitude) and the algorithm partition of the cell line panel into two groups at an expression level of 841. The *p*-value cutoff was 0.001, the *F*-test cutoff was 0.01 and the coefficient of variation was 0.3.

p-value = 0.0245	Low *CD55* Expression	High *CD55* Expression
Dense Cell Lines	**2**	**5**
Lumen Cell Lines	**8**	**1**

Figure 4.2 2 x 2 table of cell lines classified by morphology and by level of *CD55* expression. At a *p*-value of 0.0245, there are associations between low *CD55* expression and the lumen colony morphology and between high *CD55* expression and the dense colony morphology.

4.2.2 CD55 protein expression levels vary among colorectal cancer cell lines

To assess the patterns of CD55 surface expression, six cell lines were selected for detailed flow cytometry analysis, viz. C80, DLD1, HCA7, HCT116, LS180 and SW1222. The six, three dense lines (DLD1, HCA7 and HCT116) and three lumen lines (C80, LS180 and SW1222), were selected to allow for in-depth analyses across a smaller panel of cell lines exhibiting a range of CD55 expression levels. Cells were harvested from two-dimensional cultures, fluorescently stained with an anti-CD55 monoclonal antibody and assessed for the distribution of CD55 expression via flow cytometry against isotype control-stained cells.

Flow cytometry histograms of control and CD55-stained cells were generated with appropriate gates to assess the level of cellular antibody positivity (Figure 4.3). Since flow cytometry allows for cell-by-cell assessments of protein expression, a population distribution of expression can be determined, rather than the average level of expression, as Western blotting or immunofluorescence ascertain. By comparing the fluorescence distribution to the control, the proportions of positive and negative cells can be quantitatively evaluated, rather than general qualitative assessments, such as low, medium or high expression.

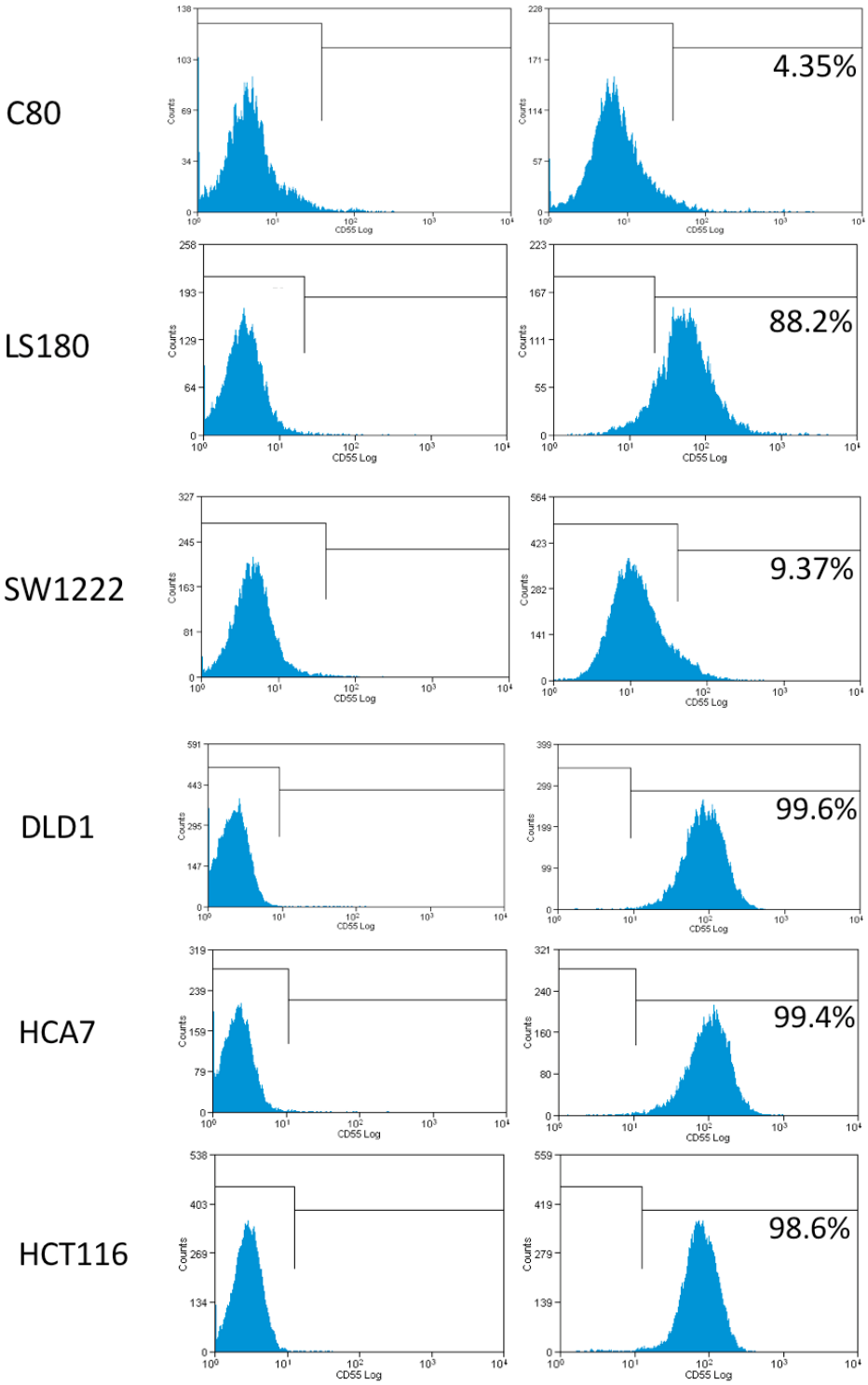


Figure 4.3 Flow cytometry histograms of CD55 expression for the C80, LS180, SW1222, DLD1, HCA7 and HCT116 cell lines. Multicellular aggregates and cell debris were excluded from the analysis, and CD55 expression was measured using the 633-nm APC filter at a voltage of 750 V. Histograms on the left are negative controls used to set the positive and negative expression gates; histograms on the right are the expression distributions. The three lumen lines express CD55 at lower levels than do the three dense lines.

The fluorescence distributions as assessed by flow cytometry were relatively symmetrical and continuous; some skewness was observed in the C80, SW1222, DLD1 and HCA7 cell lines. The existence of skewed or multimodal distributions would imply that within the population of cells existed multiple subpopulations of cells, each with their own unique distributions of CD55 expression, so for four cell lines, there appear to be at least two separate populations based on CD55 expression that could potentially be separated.

Among the lumen lines, significant subpopulations of the cells did not express CD55 at all; two lines, C80 and SW1222, were primarily CD55-negative in expression, while the remaining cell line, LS180, possessed a small minority of CD55-negative cells. Among the dense lines, all three strongly expressed CD55, with the HCT116 possessing a small minority subpopulation that did not highly express CD55. Observation of the fluorescence peaks indicated that the dense lines tended to express higher levels of CD55 than the lumen lines (the LS180 lumen line being a notable exception), and the distributions of the dense lines tended to be narrower than those of the lumen lines, indicative of a higher degree of homogeneity with respect to CD55 expression among the dense lines compared to the lumen lines. The results of the flow cytometry experiments indicate that on a protein level, dense lines express higher overall levels of CD55 than do lumen lines, an observation congruent with the results of the microarray analysis.

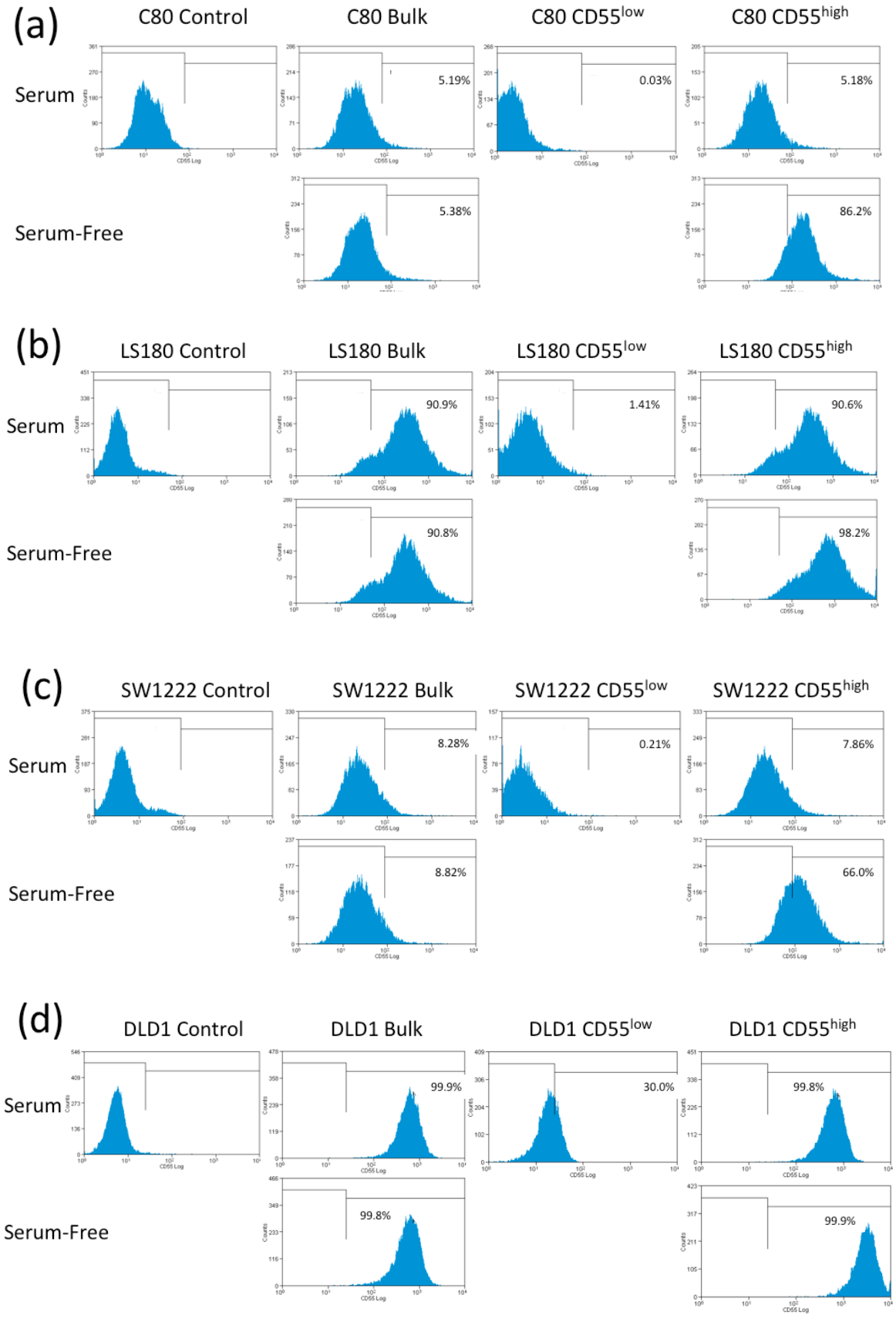
4.2.3 Fluorescence-activated cell sorting of CD55 identifies two populations of cells with different differentiation potentials

Of the six lines assessed for CD55 expression via flow cytometry, only two possessed cell populations that could be divided into clearly positive and clearly negative components. Since the difference in mRNA expression was found to be statistically significant, the effects of CD55 on the maintenance of stemness or the promotion of differentiation could

possibly be related to the specific level of protein expression. To that end, the C80, DLD1, HCA7, HCT116, LS180 and SW1222 cell lines were sorted on the basis of CD55 expression via fluorescence-activated cell sorting (FACS).

To tease out the relationship between CD55 expression and the CSC phenotype, cells were stained with anti-CD55 antibody. The resulting one-color fluorescence distribution was separated into three populations, namely the most fluorescent 1% on the histogram (the $CD55^{high}$ cells), the least fluorescence 1% on the histogram (the $CD55^{low}$ cells) and the unsorted, bulk population control. Approximately 3×10^7 live cells were sorted and the approximately 250,000 cells isolated from each of the top and bottom sectors of the histogram were divided and cultured in serum-containing medium and in defined serum-free medium. The FBS in serum-containing medium contains growth factors, chemokines and hormones thought to promote differentiation, while the defined medium contains specific quantities of salts, cofactors, hormones and growth factors (particularly EGF and bFGF) that promote stemness (Ricci-Vitiani, et al., 2007). True CSCs should retain their surface phenotype if cultured in serum-free medium; their phenotype should diverge and approach that of the original bulk culture if cultured in serum medium. Simultaneously, true differentiated cells should retain their phenotype upon culture in serum medium and be unable to give rise to any other cell surface phenotype.

Sorted cells were separately cultured in the two media for one week—cultures were subsequently reassessed for expression of CD55 via flow cytometry (Figure 4.4), and the resulting fluorescence distributions were assessed for differences (Table 4.2). In terms of cell growth, the bulk and $CD55^{high}$ populations tolerated both media and thrived, while the $CD55^{low}$ population only tolerated the serum medium; few cells survived the serum-free media, preventing an analysis of $CD55^{low}$ cells from the culture condition.


(a)
C80 Control
C80 Bulk
C80 CD55low
C80 CD55high
Serum
5.19%
0.03%
5.18%
Serum-Free
5.38%
86.2%
(b)
LS180 Control
LS180 Bulk
LS180 CD55low
LS180 CD55high
Serum
90.9%
1.41%
90.6%
Serum-Free
90.8%
98.2%
(c)
SW1222 Control
SW1222 Bulk
SW1222 CD55low
SW1222 CD55high
Serum
8.28%
0.21%
7.86%
Serum-Free
8.82%
66.0%
(d)
DLD1 Control
DLD1 Bulk
DLD1 CD55low
DLD1 CD55high
Serum
99.9%
30.0%
99.8%
Serum-Free
99.8%
99.9%

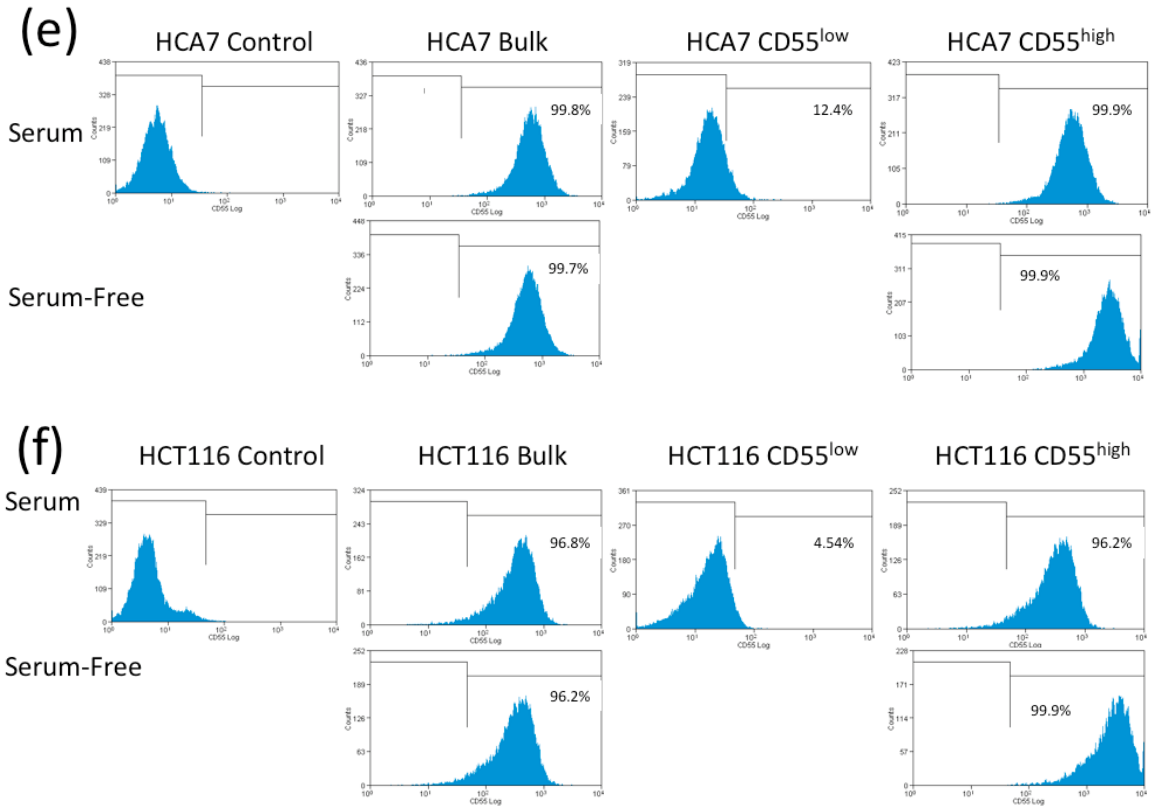


Figure 4.4 Flow cytometry histograms of CD55 expression for the C80 (a), LS180 (b), SW1222 (c), DLD1 (d), HCA7 (e) and HCT116 (f) cell lines. Sorted cells were cultured in serum or serum-free media prior to analysis. Multicellular aggregates and cell debris were excluded from the analysis, and CD55 expression was measured using the 633-nm APC filter at a voltage of 750 V. The percentage of the distribution more fluorescent than the isotype control population was quantified. Serum-free culture of $CD55^{high}$ cells revealed the ability of such cells to retain their high-CD55 phenotypes, while serum cultures revealed their abilities to differentiate and lose CD55 expression. $CD55^{low}$ cells were revealed to be unable to differentiate or to give rise to any other CD55 populations.

Cell Line	Serum-Free Media		Serum Media		
	Bulk	$CD55^{high}$	Bulk	$CD55^{low}$	$CD55^{high}$
C80	5.38%	86.2%	5.19%	0.03%	5.18%
LS180	90.8	98.2	90.9	1.41	90.6
SW1222	8.82	66.0	8.28	0.21	7.86
DLD1	99.8	99.9	99.9	30.0	99.8
HCA7	99.7	99.9	99.8	12.4	99.9
HCT116	96.2	99.9	96.8	4.54	96.2

Table 4.1 Summary of CD55-positive percentage data from differential media culture analysis. Serum-free culture of $CD55^{high}$ cells revealed the ability of such cells to retain their high positivity phenotypes, while serum cultures revealed their abilities to differentiate. $CD55^{low}$ cells were revealed to be unable to differentiate or to give rise to any other CD55 populations.

Among the cell lines, culture of $CD55^{high}$ cells in serum-containing media led to the differentiation of the cells toward a lower or more heterogeneous CD55 expression distribution comparable to that of the cultured bulk populations. The CD55-positivities of $CD55^{high}$ serum-cultured cells were similar to those of unsorted bulk cells. Among the dense cell lines, the peaks of the fluorescence distributions of the $CD55^{high}$ cells were also similar to those of the bulk population, indicating that the dense $CD55^{high}$ cells also lost a proportion of their CD55 expression as they differentiated back into a normal culture.

Culture of $CD55^{high}$ cells in serum-free media revealed a different phenotype—the cells appeared to retain their high CD55 expression profile, even after additional cell growth during the week of culture. C80 and SW1222 $CD55^{high}$ cells were far more positive for CD55 than were the bulk cultures, while comparisons of the distributions of LS180, DLD1, HCA7 and HCT116 $CD55^{high}$ cells to the unsorted cells indicated that the $CD55^{high}$ cells expressed CD55 at far higher levels than did the bulk cells. The serum-free medium allowed $CD55^{high}$ cells to retain their phenotype and not differentiate.

$CD55^{low}$ cells cultured in serum media were unable to recapitulate the CD55 expression distribution of the bulk population—the cells were far less positive for CD55 than were the bulk cells, and the peaks of the fluorescence distributions were lower than those of the bulk cells. Clearly, $CD55^{low}$ cells were already in a differentiated state (by virtue of their low CD55 expression) and were unable to respond to external cues to give rise to CD55-positive cells at levels comparable to those of the bulk culture. $CD55^{low}$ cells are indeed differentiated and are unable to exhibit plasticity of phenotype, a hallmark of CSCs.

4.2.4 $CD55^{low}$ and $CD55^{high}$ cell populations exhibit different clonogenicity propensities

Since the $CD55^{low}$ and $CD55^{high}$ cell subpopulations exhibited vastly different propensities to self-renew and differentiate, other characteristics of CSCs, including the ability to initiate a clone (the *in vitro* surrogate of an *in vivo* tumor) within a three-dimensional growth matrix. To that end, cells were stained with anti-CD55 antibody. The resulting one-color fluorescence distribution was separated into three populations, namely the more fluorescent 1% on the histogram (the $CD55^{high}$ cells), the least fluorescence 1% on the histogram (the $CD55^{low}$ cells) and the unsorted, bulk population control. Sorted cells (and unsorted, bulk culture cells) were grown in triplicate in 1,000-cell aliquots in Matrigel™ in 96-well plates (1,000 cells per well) for two weeks to assess their clonogenic abilities in three-dimensions. Hypothetically, the CSC population, in this case the $CD55^{high}$ cells, should be able to generate more colonies than the bulk culture, or any other subpopulation within the culture. After two weeks of culture, wells were photographed (Figure 4.5), and the number and type of colonies within each well was assessed (Table 4.2 and Figure 4.6).

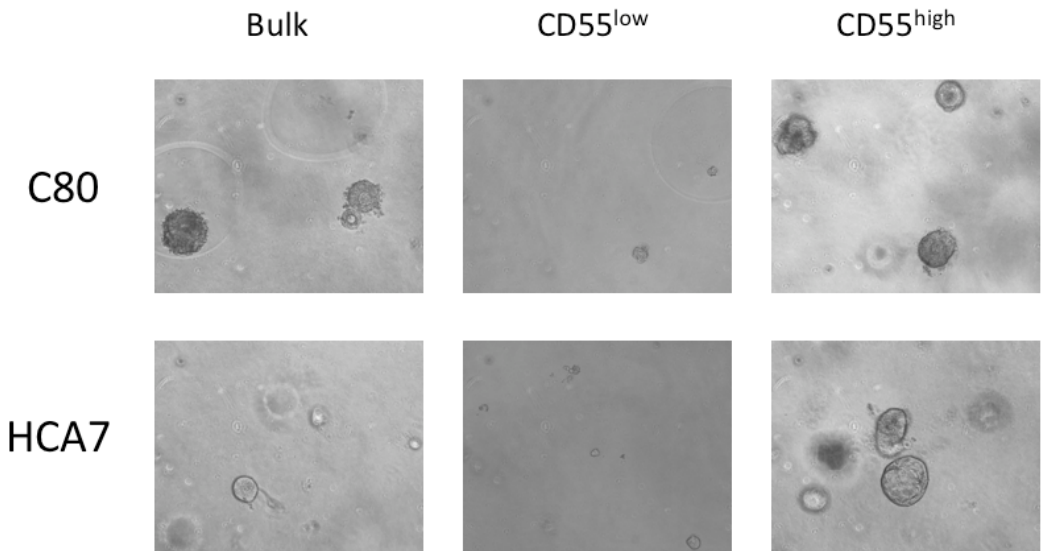


Figure 4.5 Example photographs (10x magnification under phase contrast) of 1,000 sorted cells from the C80 and HCA7 cell lines in wells of a 96-well plate. $CD55^{high}$ cells are able to initiate more colonies within Matrigel™ than the bulk culture, while $CD55^{low}$ cells are able to initiate few, if any colonies.

Cell Line	Bulk		$CD55^{low}$		$CD55^{high}$	
	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens
C80	38 ± 3.9	0.36 ± 0.03	11 ± 2.2	0.09 ± 0.04	88 ± 4.7	0.48 ± 0.02
LS180	41 ± 4.7	0.29 ± 0.01	10 ± 2.6	0.08 ± 0.04	102 ± 7.0	0.41 ± 0.02
SW1222	25 ± 3.8	0.32 ± 0.01	5 ± 1.2	0.05 ± 0.05	62 ± 4.5	0.45 ± 0.03
DLD1	68 ± 4.7	0.00	10 ± 2.6	0.00	95 ± 8.1	0.00
HCA7	64 ± 4.4	0.00	11 ± 2.2	0.00	83 ± 4.7	0.00
HCT116	80 ± 4.3	0.00	16 ± 4.6	0.00	123 ± 7.6	0.00

Table 4.2 Summary of clonogenicity data from fluorescence-activated cell sorting on the basis of CD55. Each value represents six data points (two experiments done in triplicate). Among the lumen lines, selection of the $CD55^{high}$ population yields higher clonogenicities and a greater proportion of colonies possessing lumens, while selection of the $CD55^{low}$ population dampens clonogenicity and lumen-formation. Among the dense lines, CD55 also enhances clonogenicity, while lack of CD55 dampens clonogenicity.

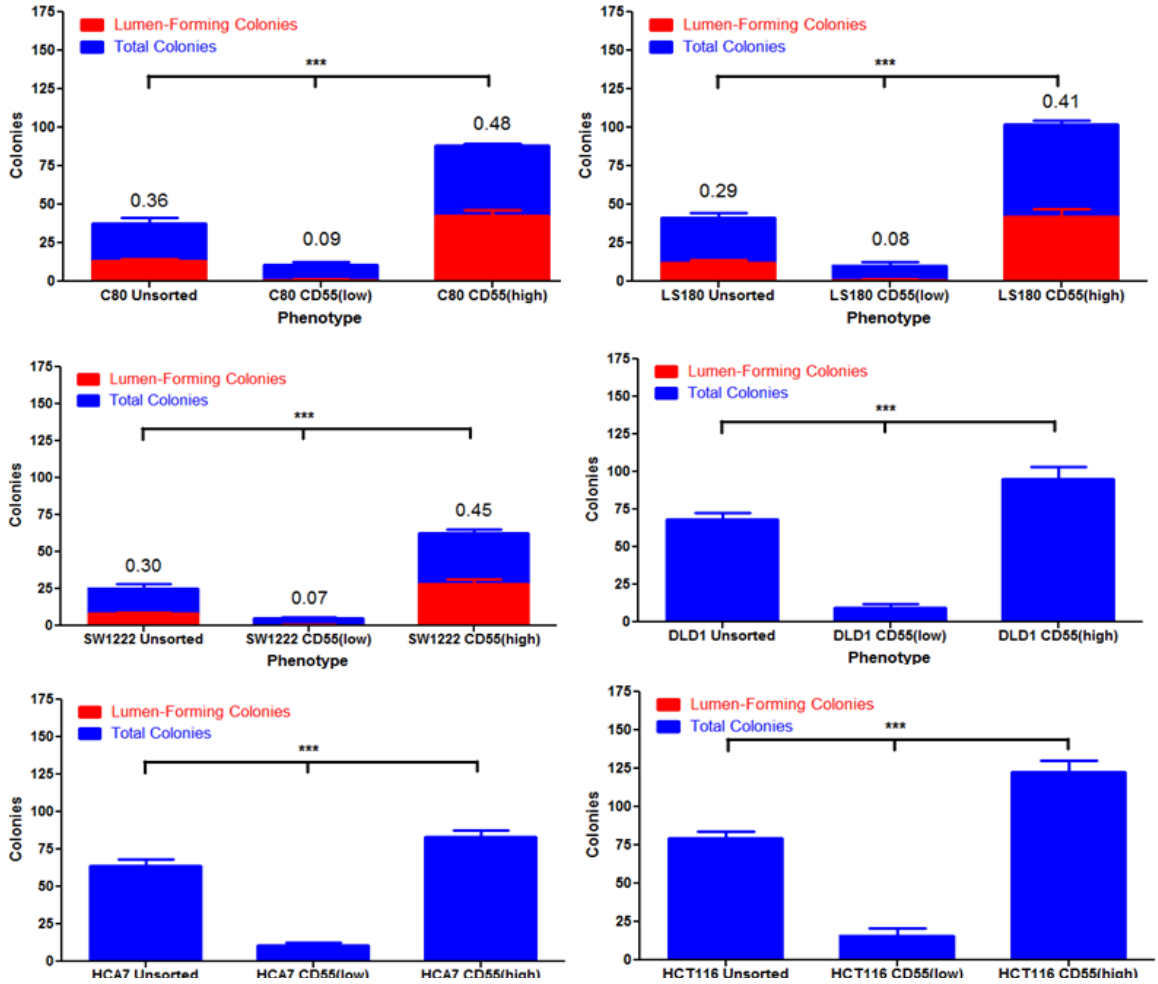


Figure 4.6 Bar graphs showing the number and morphology of colonies (from Table 4.2) generated by sorted cells. Cell lines were sorted on the basis of CD55 expression, divided into 1,000-cell aliquots and grown in Matrigel™ for two weeks. Each bar represents six data points (two experiments done in triplicate). Among the lumen lines C80, LS180 and SW1222, selection of the $CD55^{high}$ population yields higher clonogenicities and a greater proportion of colonies possessing lumens, while selection of the $CD55^{low}$ population dampens clonogenicity and lumen-formation. Among the dense lines DLD1, HCA7 and HCT116, CD55 also enhances clonogenicity, while lack of CD55 dampens clonogenicity. Mean proportions of lumens and mean clonogenicities were compared by one-way analyses of variance (ANOVAs); both lumen proportion and clonogenicity differences were significant at a 0.005 significance level for all six cell lines. *** - 0.005 significance level.

The lumen cell lines show a significant increase in clonogenicity when the $CD55^{high}$ cells were purified and, interestingly, an increase in the proportion of colonies that possess lumen-like structures. The C80 $CD55^{high}$ cells gives rise to more than double the colonies

than did the bulk culture, while LS180 and SW1222 $CD55^{high}$ cells generated more than triple the number of colonies, each. Furthermore, $CD55^{high}$ cells generated a 33% increase in the proportion of colonies that contain lumen-like structures. Among the dense cell lines, increases were also observed in clonogenicity of the $CD55^{high}$ cells, although the increases for DLD1, HCA7 and HCT116 were approximately 50% rather than the 100% increases seen by the lumen cell lines. Furthermore, no nascent lumen formation was observed among the $CD55^{high}$ dense cell line cells.

The $CD55^{low}$ cells, almost uniformly, were unable to general colonies at numbers comparable to the bulk cultures. Clonogenicity was reduced to approximately 25% of the original level, while the capacity to form lumens also fell precipitously to less than 25% of the original levels among the lumen lines.

Clearly, the $CD55^{high}$ cell population is able to generate more colonies *in vitro* than both the $CD55^{low}$ cell population and the bulk population; all of the cell lines showed increased clonogenicity for the $CD55^{high}$ population at a significance level of 0.005 as compared to the other two populations. Indeed, the bulk population clonogenicity appears to be an average of the clonogenicities of the $CD55^{low}$ and $CD55^{high}$ subpopulations. Furthermore, $CD55^{high}$ cells are able to form more colonies that possess lumen-like structures. Since the formation of lumens is connected to the ability to differentiate into the lineages of the colon, perhaps the increase in lumen-containing colonies among $CD55^{high}$ cells is an indication that $CD55^{high}$ cells represent a more pure population of CSCs that are able to differentiate. The dampening of the ability to form lumens among $CD55^{low}$ cells could potentially be linked to their differentiated status—differentiated cells would be unable to form the luminal structures because of their lack of plasticity. Thus, on the basis of clonogenicity and lumen-formation, CD55 appears to be linked to the CSC phenotype.

4.2.5 $CD55^{low}$ and $CD55^{high}$ cell populations exhibit different proliferative propensities

In addition to possessing the abilities of self-renewal, plasticity and tumor initiation, CSCs must also possess the ability to drive the growth of a tumor—otherwise, CSC-initiated tumors would remain small, relatively benign and insignificant. Culturing $CD55^{high}$ cells in serum-free media demonstrated self-renewal, culturing $CD55^{high}$ cells in serum media demonstrated plasticity and culturing $CD55^{high}$ cells in Matrigel™ demonstrated tumor initiation. To that end, cells were stained with anti-CD55 antibody. The resulting one-color fluorescence distribution was separated into three populations, namely the most fluorescent 1% on the histogram (the $CD55^{high}$ cells), the least fluorescence 1% on the histogram (the $CD55^{low}$ cells) and the unsorted, bulk population control. Sorted cells (and unsorted, bulk culture cells) were grown in triplicate in 3,000-cell aliquots in 96-well plates for one week. Every two days, the level of cell proliferation was assessed by the MTS colorimetric assay. The results of the assay were plotted as growth curves (Figure 4.7), and the slopes of the curves were compared to ascertain the subpopulation with the greatest proliferative potential (Table 4.3). Hypothetically, the CSC population should be able to grow more quickly than the bulk culture or the non-CSC population.

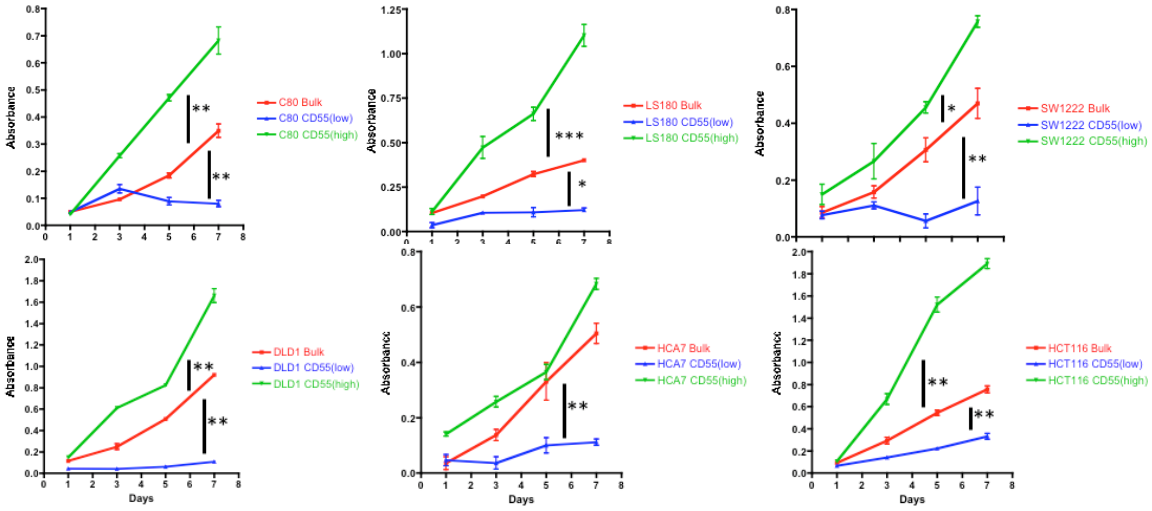


Figure 4.7 Growth curves corresponding to MTS-measured cell proliferation for sorted cell fractions on the basis of CD55 expression. Plotted points represent six data points (two experiments done in triplicate). For all cell lines tested, the $CD55^{high}$ fraction was the most proliferative and grew the most rapidly over a seven-day test period, while the $CD55^{low}$ fraction was the least proliferative. The slopes of the sorted populations were compared to that of the unsorted, bulk, populations via Student's *t*-tests. * - 0.05 significance level, ** - 0.01 significance level, *** - 0.005 significance level.

Cell Line	Bulk Culture Slope (units/day)	$CD55^{low}$ Slope (units/day)	$CD55^{high}$ Slope (units/day)
C80	0.0493 ± 0.0095	0.0024 ± 0.0010	0.1069 ± 0.0004
LS180	0.0506 ± 0.0030	0.0130 ± 0.0054	0.1579 ± 0.0162
SW1222	0.0650 ± 0.0076	0.0046 ± 0.0080	0.1006 ± 0.0147
DLD1	0.1335 ± 0.0223	0.0108 ± 0.0038	0.2374 ± 0.0427
HCA7	0.0800 ± 0.0070	0.0129 ± 0.0048	0.0867 ± 0.0178
HCT116	0.1122 ± 0.0033	0.0440 ± 0.0028	0.3105 ± 0.0313

Table 4.3 Summary of proliferation data from FACS on the basis of CD55. Presented values represent six data points (two experiments done in triplicate). Among the lumen lines, selection of the $CD55^{high}$ population increased growth rate relative to the bulk culture, while selection of the $CD55^{low}$ population reduced proliferation. Among the dense lines, the results are similar, but the differences between the proliferation rates of the $CD55^{high}$ cells and the bulk cells were less stark than in the lumen lines.

The lumen cell lines C80, LS180 and SW1222 grew relatively slowly over the week compared to the dense cell lines—the growth rates of the lumen lines were 50-75% of the growth rates of the dense lines, an observation in line with previous assessments of dense line clonogenicity. When the $CD55^{high}$ subpopulation was purified, cell proliferation was enhanced two-to-three-fold, with LS180 showing the greatest increase and SW1222 the smallest. Since the LS180 cell line was the most positive for CD55 among the three lumen lines, it is puzzling that the greatest difference between the bulk culture and the $CD55^{high}$ culture was observed within the LS180 cell line. Perhaps the LS180 CD55-positive population contains distinct subpopulations and the $CD55^{high}$ population corresponds to the most CSC-like subpopulation, or there is another unknown mechanism.

Among the dense lines, when the $CD55^{high}$ subpopulation was purified, cell proliferation generally was enhanced two-to-three-fold as well, with HCT116 showing the greatest increase in proliferative ability and HCA7 showing the least increase. Indeed, HCA7 $CD55^{high}$ cells grew at a rate only slightly higher than that of the bulk culture. Perhaps for HCA7, CD55 does not subdivide the population into groups of radically-different growth potentials as it does for the other two dense cell lines. In any case, $CD55^{high}$ cells were more proliferative than the bulk cultures for all six tested cell lines.

The $CD55^{low}$ cells, uniformly, were unable to grow at any appreciable rate. All six $CD55^{low}$ cultures saw their growth rates drop to 10-20% of their respective bulk culture growth rates. Even at these low rates, an approximation of the mean growth rate of the $CD55^{low}$ and $CD55^{high}$ cells approaches that of the bulk culture, indicating that $CD55^{low}$ and $CD55^{high}$ cells represent different subpopulations of differing proliferative abilities that, when combined, are able to grow at the rate of the bulk culture.

Based on these results, $CD55^{high}$ cells appear to be a subpopulation within bulk cultures that are able to sustain the overall growth of the cell line, as demonstrated by their high growth rates relative to the other two populations assessed. Clearly, the $CD55^{high}$ cells behave like tumor-driving CSCs, while $CD55^{low}$ cells behave like differentiated cells.

4.2.6 $CD55^{low}$ and $CD55^{high}$ cell populations exhibit different cell cycle profiles

There is evidence to suggest that CSCs exhibit an altered cell cycle profile. Harper and colleagues (2010) isolated $CD44^{high}$ and $CD44^{low}$ populations from a panel of seven epithelial cancer cell lines (head and neck, prostate and breast cancers) and eleven epithelial cancer primary tissue samples (head and neck only) and assessed their DNA content via flow cytometry. Peaks in the resulting histogram corresponding to the G_1, S and G_2 phases of the cell cycle were analyzed to ascertain the distribution of cells within those respective phases. $CD44^{low}$ differentiated cells were found to possess a greater proportion of cells within the G_1 phase and $CD44^{high}$ cancer stem cells were found to possess a greater proportion of cells within the G_2 phase, relative to the unsorted bulk population. It is likely that the preponderance of the G_1 phase in $CD44^{low}$ cells is attributable to their inability to proliferate and their differentiated phenotype, while the G_2 block exhibited by the $CD44^{high}$ may lead to the activation of checkpoint kinases that confer a proliferative and tumor-initiating phenotype upon cells (Chappell and Dalton, 2010).

To assess if $CD55^{low}$ and $CD55^{high}$ cells also exhibited cell cycle perturbations consistent with their classifications as cancer differentiated and stem cells, respectively, cell lines were stained with anti-CD55 antibody and sorted via FACS. The resulting one-color fluorescence distribution was separated into three populations, namely the most fluorescent 1% on the histogram (the $CD55^{high}$ cells), the least fluorescence 1% on the

histogram (the $CD55^{low}$ cells) and the unsorted, bulk population control. Approximately 3 x 10^7 live cells were sorted and the approximately 250,000 cells isolated from each of the top and bottom sectors of the histogram were fixed with ethanol, stained with Hoechst 33342 (a DNA intercalating dye) and analyzed via flow cytometry (Figure 4.8). The resulting histograms were used to ascertain the proportions of the distribution within the G_1, S and G_2/M phases (Figure 4.9). Gates were drawn around the specific peaks and troughs of the distribution corresponding to the level of DNA staining, with the G_1 peak corresponding to the tallest peak, the G_2/M peak corresponding to the peak of approximately double the fluorescence of the G_1 peak and the S trough corresponding to the range of DNA staining fluorescences between G_1 and G_2/M.

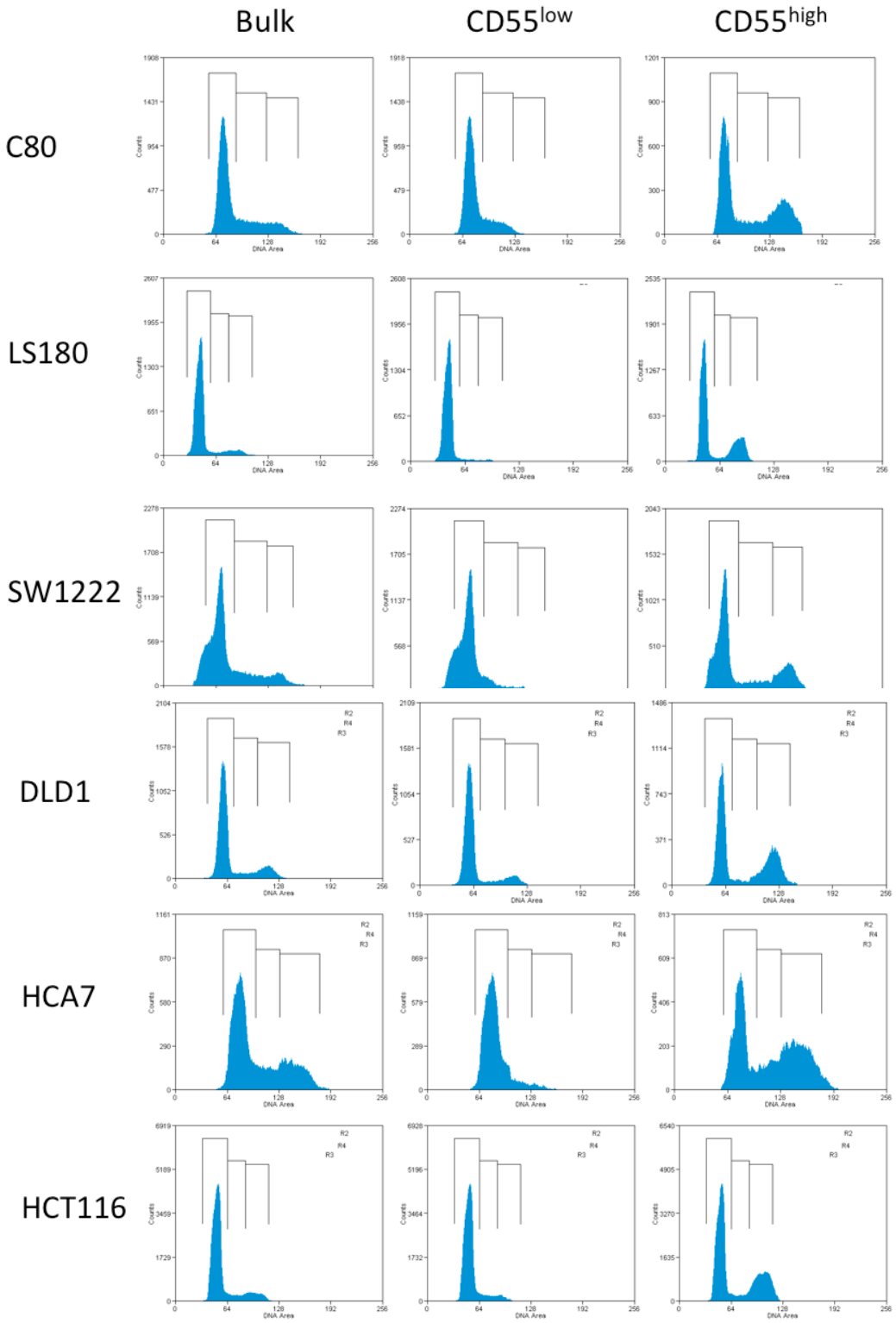


Figure 4.8 Cell cycle profiles of cells sorted on the basis of CD55 expression. Cell fractions were ethanol-fixed and stained with the DNA dye Hoechst 33342 to assess cellular DNA content and, therefore, the distribution of cell cycle phases. Gates were drawn to divide the histogram into the constituent three cell cycle phases. A distinct trend toward the G_1 phase for $CD55^{low}$ cells and toward the G_2/M phase for $CD55^{high}$ cells was noticed.

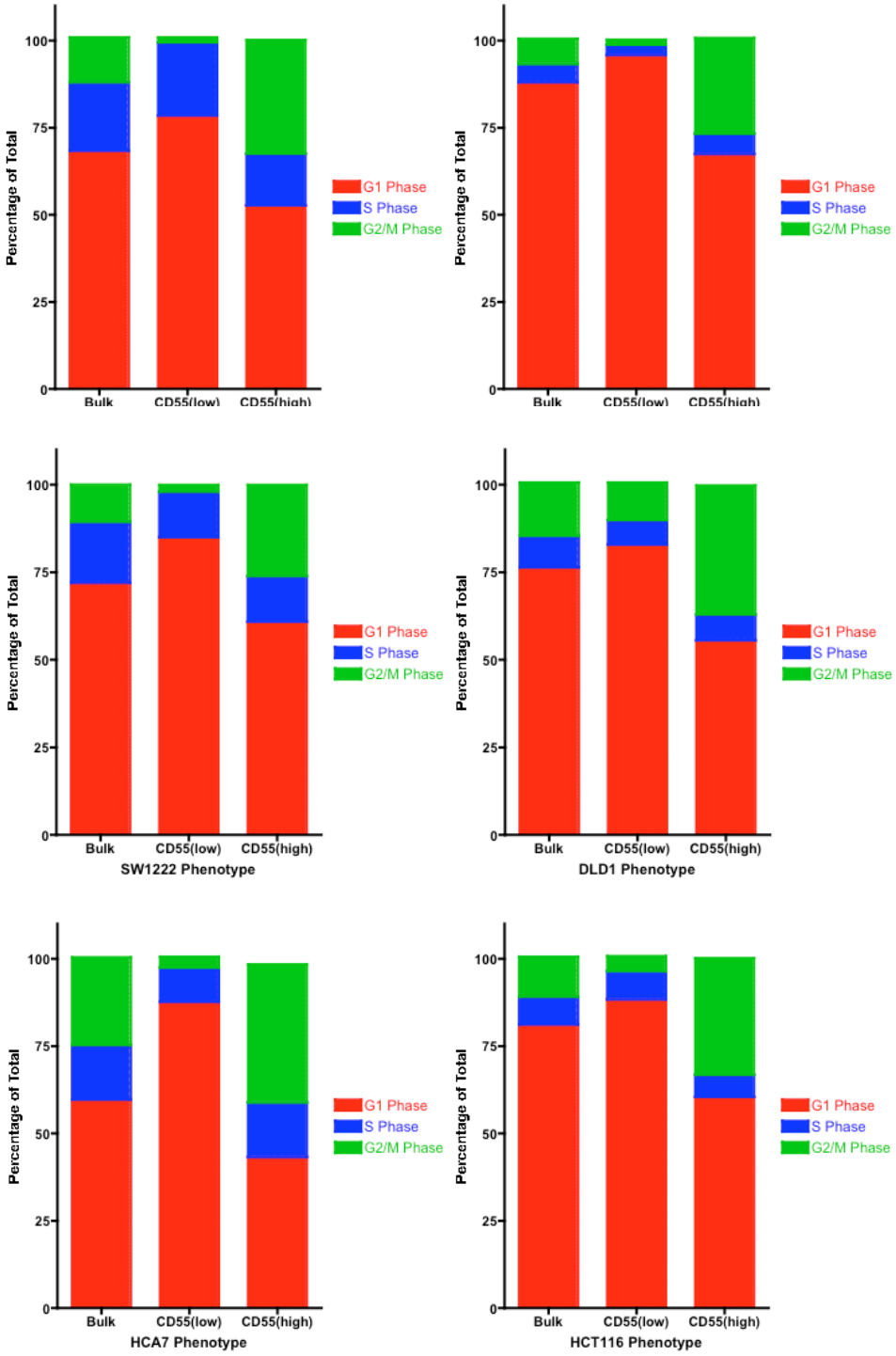


Figure 4.9 Summary of cell cycle phase distributions from sorted cell fractions on the basis of CD55 expression. The $CD55^{low}$ subpopulation is associated with an increased population of G_1-phase cells, while the $CD55^{high}$ subpopulation is associated with an increased population of G_2/M-phase cells, all relative to the bulk population.

The cell cycle profiles of the bulk cultures show no distinct bias toward either the G_1 or G_2/M phases. For example, the lumen line LS180 has a large proportion of its cells residing in the G_1 phase, while the same is true for the HCT116 dense line. Conversely, the dense line HCA7 and the lumen line C80 both have large minorities of their cells residing in the G_2/M phases. The G_1 phase in general occupied 60-90% of all the cells for all six lines, but in each cell line, the profile is unique.

Among the lumen lines, selection of $CD55^{high}$ cells led to a small general decrease in the G_1 phase population of approximately 10%; the G_1 peaks among the $CD55^{high}$ cells were very similar, *prima facie*, to those of the bulk, unsorted cultures. A vastly different result was obtained from an analysis of the G_2/M peaks, with $CD55^{high}$ cells showing a three-to-four-fold increase in G_2/M population versus the bulk cultures. Interestingly, the LS180 line showed the largest relative increase in the G_2/M peak, a four-fold increase, and the largest relative decrease in the G_1 peak. Since LS180 expresses the highest level of CD55 among the three lumen lines tested, and its CD55 flow cytometry distribution shows most of its cells to be positive for CD55, it is paradoxical that such stark changes in the cell cycle profile between the bulk culture and the $CD55^{high}$ culture are observed. It is likely, therefore, that within the positive CD55 subpopulation exist subsidiary cell fractions with different behaviors and phenotypes. Selection of the $CD55^{low}$ cells yielded opposite results; the G_1 peak rose by approximately 10% for all three lines, while the G_2/M peaks withered to 10-20% of the bulk cultures' frequencies. Again, LS180 stands out as an anomaly among the lumen lines, in that its $CD55^{low}$ population has its G_2/M peak shorten by 80%, rather than the 90% exhibited by C80 and SW1222.

Among the dense lines, selection of $CD55^{high}$ cells also led to a small general decrease in the G_1 phase population of approximately 10%; the G_1 peaks among the $CD55^{high}$ cells

were also similar to those of the bulk cultures. At the same time, $CD55^{high}$ cells showed increases in G_2/M population versus the bulk cultures, but on the order of two-to-three fold rather than the three-to-four-fold exhibited by the lumen lines. Since all three dense lines are strongly positive for CD55, and thus their respective bulk cultures express CD55 almost ubiquitously, it is fascinating to observe intra-distributional effects of CD55, in that specific, high, levels of CD55 can be linked to changes in the cell cycle profile and to an accumulation of cells within the G_2/M phase. Selection of the $CD55^{low}$ cells yielded reverse results; the G_1 peak rose by approximately 10% for two lines (HCA7 had its G_1 peak rise by 20%). However, the G_2/M peak only fell dramatically within the HCA7 line; for DLD1 and HCT116, the fall was only 33-50%. Here, HCA7 stands out as a cell line for which the $CD55^{low}$ population is very different from the bulk culture.

4.2.7 $CD55^{low}$ and $CD55^{high}$ cell populations express varying levels of CDX1 and BMI1

Within the context of these experiments, CD55 is regarded as a novel marker of cancer stemness; thus, an examination of its properties would initially focus on its abilities to select for a subpopulation of cells that are able to self-renew, differentiate, initiate clones and drive overall cell proliferation. A more detailed examination of these putative CSCs would involve probing for activation of cellular signaling pathways known to be involved either in the maintenance of a stem-like phenotype or in the suppression of differentiation.

As previously stated, CDX1 is a transcription factor linked to differentiation in the gut epithelium. Immunohistochemical staining of colonic crypts has revealed CDX1 expression to be high near the luminal surface, where the differentiated cell population resides, and low near the crypt base, where the stem cell population is thought to reside

(Chan, et al., 2009). Furthermore, CDX1 expression dampens proliferation and tumorigenicity (Mallo, et al., 1998), two behaviors associated with cancer stem cells. BMI1, on the other hand, is a member of the Polycomb group (PcG) of intracellular proteins; BMI1 associates with other PcG proteins to form the multiprotein complex Polycomb Repressive Complex 1 (PRC1) to direct chromatin modifications that repress the transcription of multiple target genes, many of which are involved in promoting differentiation (Gil, et al., 2005). These stemness-promoting properties of BMI1 can be observed within colorectal cancer: an immunohistochemical analysis of colorectal tissue observed that only 7.9% of normal bowel tissue stained positive for BMI1, while 66.5% and 86.4% of colorectal tumors and lymph node metastases, respectively, stained positive (Li, et al., 2010). Furthermore, the aforementioned colony morphology microarray analysis linked *BMI1* mRNA expression to the undifferentiated dense colony morphology. Based upon these previous findings, it is logical to hypothesize that a cancer stem cell subpopulation would express lower levels of CDX1 and higher levels of BMI1 relative to the bulk cell population.

To test this hypothesis within the context of CD55, cells were sorted on the basis of CD55 expression via FACS and lysed with RIPA buffer. Resulting cellular lysates were electrophoresed, transferred to a nitrocellulose membrane and probed for expression of CDX1 and BMI1 with monoclonal antibodies via Western blotting (Figure 4.10). A monoclonal antibody against β-tubulin was used as a loading control to establish that differences in protein expression were not attributable to variations in the amount of total protein loaded.

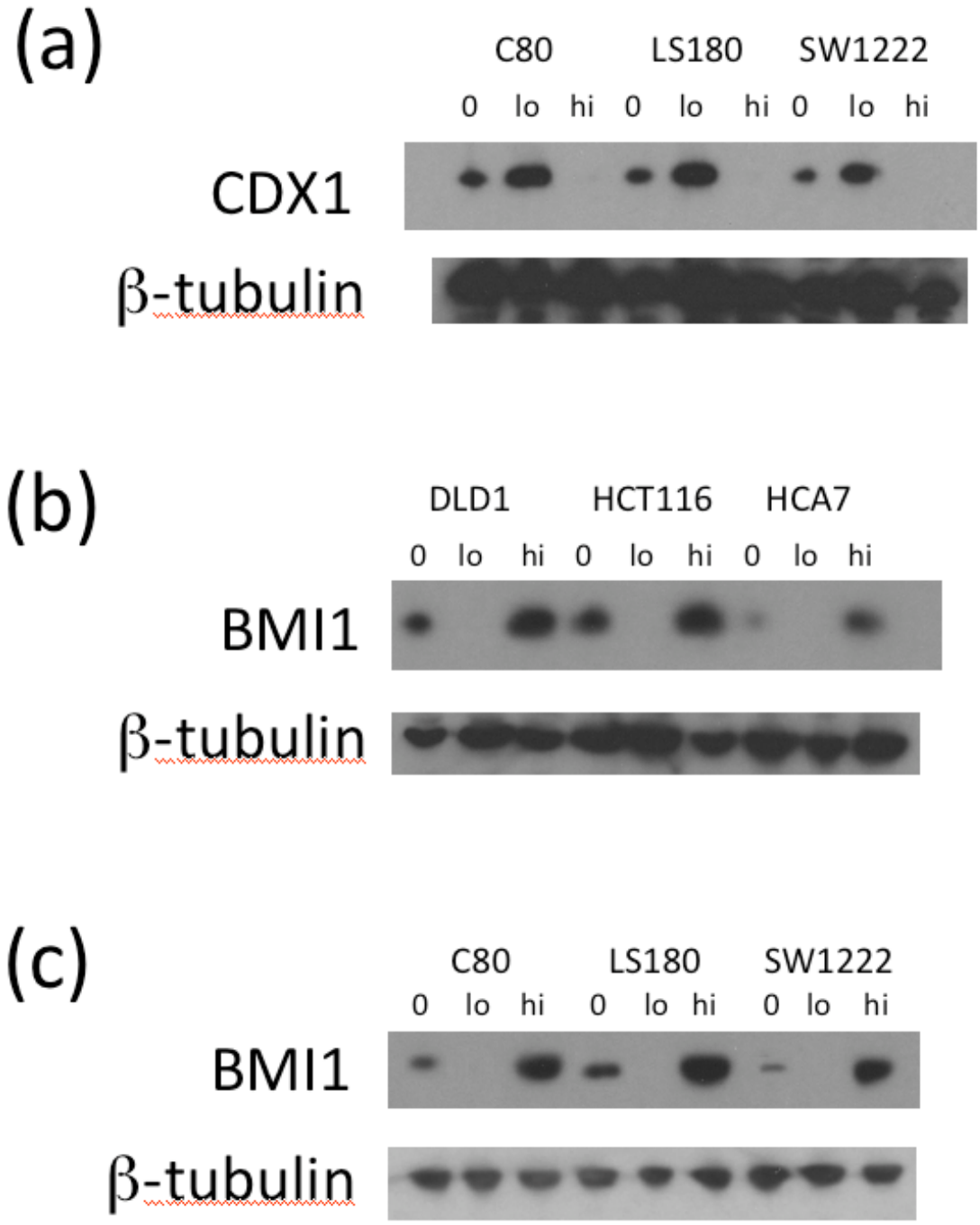


Figure 4.10 Western blots showing varying levels of expression of CDX1 within lumen lines (a) and of BMI1 within dense (b) and lumen (c) lines separated into the bulk (0), $CD55^{low}$ (lo) and $CD55^{high}$ (hi) subpopulations. Since dense lines did not express CDX1, only sorted cells from the lumen lines were probed for CDX1. Across all six cell lines, the $CD55^{high}$ population was associated with an increased expression of BMI1 and a decreased expression of CDX1.

Among the lumen cell lines assessed, Western blotting of sorted cell population lysates for CDX1 expression revealed a negative relationship between CD55 expression and CDX1 expression. Across all three lines, the $CD55^{high}$ populations expressed little or no CDX1, while the bulk and $CD55^{low}$ populations expressed CDX1. Dense cell line lysates were also probed for CDX1, but no CDX1 expression was detected within any cell fraction for DLD1, HCA7 and HCT116, in line with microarray data relating to *CDX1* mRNA expression. Therefore, the resulting (blank) Western blots are not presented.

Among the dense cell lines assessed, Western blotting of sorted cell population lysates for BMI1 expression revealed a positive relationship between CD55 expression and BMI1 expression. Across all three lines, the $CD55^{high}$ subpopulations expressed higher levels of BMI1 than did the bulk populations, with the greatest difference in expression levels found between the HCA7 $CD55^{high}$ subpopulation and the bulk population. Within the HCT116 cell line, the difference was only slight. The $CD55^{low}$ populations, interestingly, expressed very little BMI1. The same relationship was found among the lumen cell lines. Since the bulk populations of the lumen lines expressed less endogenous BMI1, the difference in expression between the $CD55^{high}$ and bulk fractions was greater than that observed in the dense cell lines. Clearly, the $CD55^{low}$ cells appear to be associated with the differentiated, CDX1-high phenotype, while the $CD55^{high}$ cells appear to be associated with the stem-like, BMI1-high phenotype, consistent with the mRNA microarray expression results linking *CDX1* expression to the differentiated lumen lines and *BMI1* expression to the undifferentiated dense lines.

4.2.8 CD24 and CD44 protein expression levels vary among colorectal cancer cell lines

Previous work in colorectal cancer cell lines has implicated the membrane proteins CD24 and CD44 with the CSC population (Yeung, et al., 2010). To assess their expression within the six-cell panel studied here (C80, LS180, SW1222, DLD1, HCA7 and HCT116), cells were harvested from two-dimensional culture, fluorescently stained with anti-CD24 and anti-CD44 monoclonal antibodies and assessed for the distributions of CD24 and CD44 expression via flow cytometry against isotype- control-stained cells.

Flow cytometry dotplots of control and stained cells were generated with appropriate gates to assess the level of cellular antibody positivity (Figure 4.11). Since flow cytometry allows for cell-by-cell quantitative assessments of protein expression, a population distribution of expression can be determined, rather than the average level of expression, as Western blotting identifies.

The dotplots reveal that there is significant diversity in the strength and distribution of CD24 and CD44 expression, with cell lines that are strongly positive for both (LS180 and HCA7), for CD24 only (SW1222), for CD44 only (HCT116) and for a range of CD24/CD44 combinations (C80 and DLD1). For most of the cell lines, subpopulations exist that are positive or negative for CD24 and CD44, thus making downstream uses of the cell lines for flow cytometry advantageous.

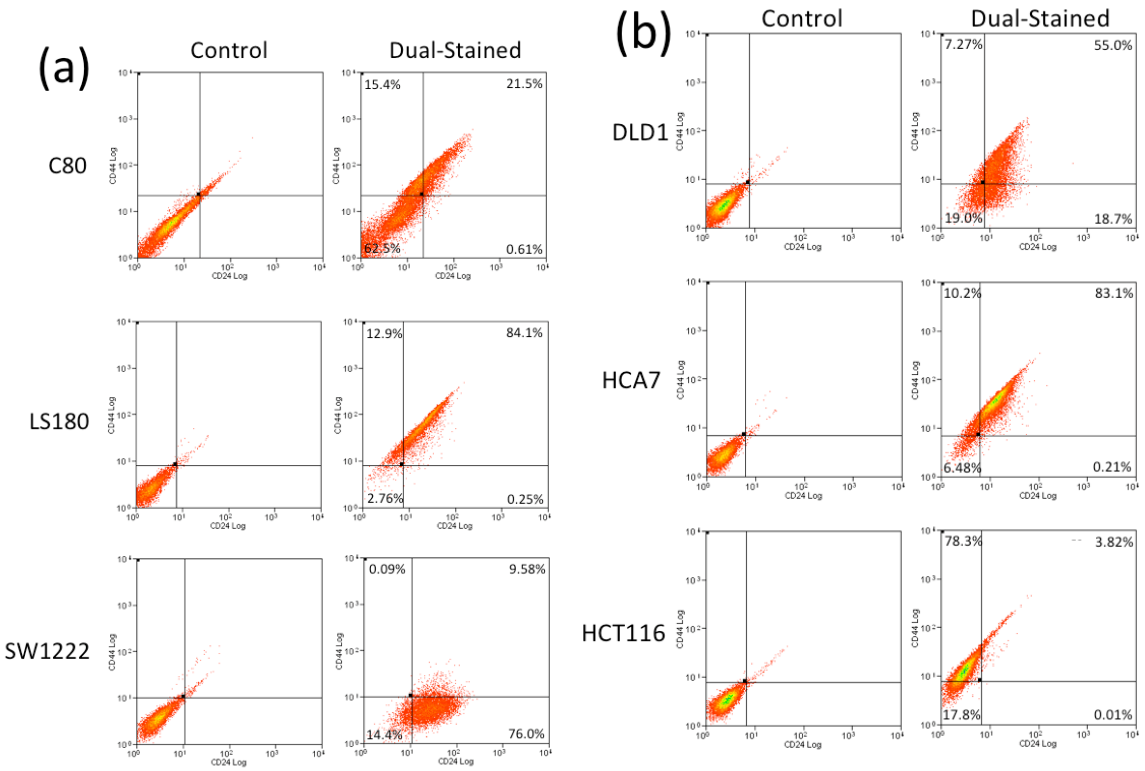


Figure 4.11 Flow cytometry dotplots of CD24 and CD44 expression for the lumen (a) and dense (b) cell lines. Multicellular aggregates and cell debris were excluded from the analysis, and CD24 and CD44 expression levels were measured using the 488-nm FITC and 488-nm PE filters at voltages of 500 V. Dotplots on the left are negative controls used to set the positive and negative expression gates. Even after flow cytometry compensation, some fluorescence spillovers are observed in some of the dotplots—the selection of the extremes of the distributions, as described later, would avoid some of these effects.

4.2.9 Adding CD55 to a CD24/CD44 dual flow cytometry sort further enhances clonogenicity

The $CD24^{high}/CD44^{high}$ cell subpopulation has been found to be associated with the cancer stem cell phenotype in colorectal (Yeung, et al., 2010) and pancreatic (Huang, et al., 2008) cancers. The evidence presented here suggests that CD55 may also be a marker of CSCs. It could be possible to combine the conclusions of these two separate bodies of work and devise a methodology for sorting cells on the basis of three markers, CD24,

CD44 and CD55, simultaneously. To that end, cell lines were sorted on the basis of CD24, CD44 and CD55 expression and separated into the most and least fluorescent 1% of the total population (the $CD24^{low}CD44^{low}CD55^{low}$ and the $CD24^{high}CD44^{high}CD55^{high}$ populations) per Figure 2.1. Sorted and unsorted cells were grown in triplicate in 1,000-cell aliquots in Matrigel™ in 96-well plates for two weeks to assess their clonogenic abilities in three-dimensions. Hypothetically, the putative CSC population, in this case the $CD24^{high}CD44^{high}CD55^{high}$ cells, should be able to generate more colonies than the bulk culture, or any other subpopulation within the culture. After two weeks of culture, the number and type of colonies within each well was assessed (Table 4.4 and Figure 4.12).

Cell Line	$CD24^{low}$ $CD44^{low}$		$CD24^{high}$ $CD44^{high}$		$CD24^{low}$ $CD44^{low}$ $CD55^{low}$		$CD24^{high}$ $CD44^{high}$ $CD55^{high}$	
	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens
C80	10 ± 1.2	0.06 ± 0.03	100 ± 3.2	0.49 ± 0.03	3 ± 0.9	0.00 ± 0.00	160 ± 4.7	0.68 ± 0.03
LS180	7 ± 1.2	0.09 ± 0.05	102 ± 4.3	0.38 ± 0.02	1 ± 0.3	0.00 ± 0.00	162 ± 5.9	0.66 ± 0.04
SW1222	4 ± 0.9	0.07 ± 0.07	75 ± 3.2	0.48 ± 0.03	0 ± 0.0	0.00 ± 0.00	134 ± 5.5	0.74 ± 0.03
DLD1	10 ± 1.5	0.00	93 ± 4.6	0.00	0 ± 0.0	0.00	164 ± 6.9	0.00
HCA7	8 ± 1.2	0.00	118 ± 6.6	0.00	3 ± 0.9	0.00	153 ± 4.6	0.00
HCT116	11 ± 2.3	0.00	141 ± 6.6	0.00	4 ± 0.8	0.00	219 ± 9.7	0.00

Table 4.4 Summary of clonogenicity data, in colonies per 1,000 cells plated, from FACS on the basis of CD24, CD44 and CD55. Presented values represent six data points (two experiments done in triplicate). Among the lumen lines, adding CD55 to the CD24/CD44-sorted population yields higher clonogenicities and a greater proportion of colonies possessing lumens, while the existence of the $CD55^{low}$ population within the dual-sorted population dampens clonogenicity and lumen-formation. Among the dense lines, CD55 also enhances clonogenicity, while lack of CD55 dampens clonogenicity.

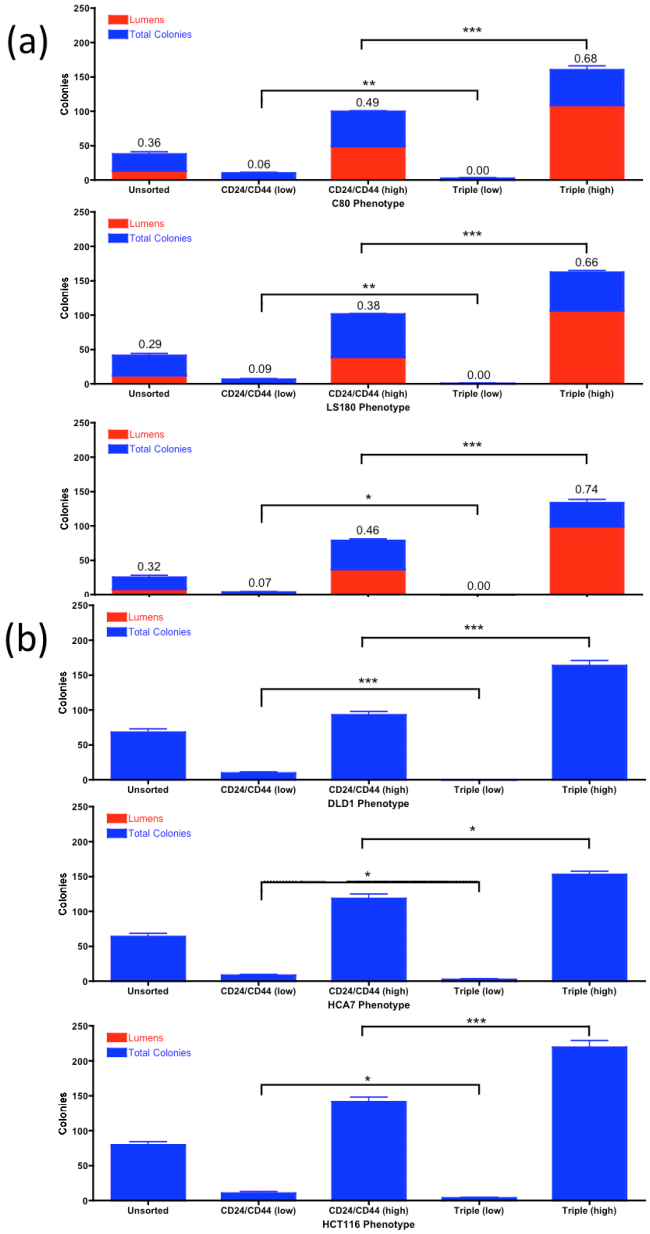


Figure 4.12 Bar graphs showing the number and morphology of colonies (from Table 4.4) generated by sorted lumen cell line cells (a) and dense cell line cells (b). Bars represent six data points (two experiments done in triplicate). Cell lines were sorted on the basis of CD24, CD44 and CD55 expression, divided into 1,000-cell aliquots and grown in Matrigel™ for two weeks. Among the lumen lines, selection of the $CD24^{high}CD44^{high}CD55^{high}$ populations yields higher clonogenicities and a greater proportion of colonies possessing lumens than the $CD24^{high}CD44^{high}$ populations, while selection of the $CD24^{low}CD44^{low}CD55^{low}$ population dampens clonogenicity and lumen-formation over the $CD24^{low}CD44^{low}$ population. Among the dense lines, CD55 also enhances clonogenicity, while lack of CD55 dampens clonogenicity. The mean proportions of lumens were compared to the unsorted, bulk population via Student's *t*-tests; all were significant at a 0.01 significance level. The mean numbers of colonies were compared to the unsorted, bulk population via Student's *t*-tests (brackets shown on graph). * - 0.05 significance level, ** - 0.01 significance level, *** - 0.005 significance level.

The lumen cell lines showed large increases in clonogenicity when CD55 was added to the $CD24^{high}CD44^{high}$ population and, as noticed with the CD55 single-marker cell sorts, showed large increases in the proportions of colonies that possess lumen-like structures. The SW1222 $CD24^{high}CD44^{high}$ $CD55^{high}$ cells gave rise to almost double the number of colonies than did the $CD24^{high}CD44^{high}$ cells, while C80 and LS180 gave rise to 60% more colonies. Clone formation and the proportion of lumens increased markedly (nearly 50%) with the addition of CD55 to the cell sort. Among the dense cell lines, increases were also observed in clonogenicity of the $CD24^{high}CD44^{high}$ $CD55^{high}$ cells over the $CD24^{high}CD44^{high}$ cells, with similar percentage increases (50%) to those of the lumen cell lines. Also, the addition of CD55 did not stimulate any lumen formation—the dense cell lines remained unable to differentiate *in vitro*.

Addition of low CD55 expression to the $CD24^{low}CD44^{low}$ populations almost completely ablated colony formation—two cell lines, SW1222 and DLD1, were unable to form colonies at all. Furthermore, among the lumen lines, propensity to give rise to lumens was completed abolished in the $CD24^{low}CD44^{low}CD55^{low}$ populations.

The $CD24^{high}CD44^{high}$ cell population was already shown to be enriched for cancer stem cells. The results here indicate that the addition of another marker, CD55, can stimulate the formation of colonies *in vitro*: all six cell lines showed increased clonogenicity for the $CD24^{high}CD44^{high}$ $CD55^{high}$ population over the $CD24^{high}CD44^{high}$ population at a significance level of 0.05, with many at 0.005. Furthermore, $CD24^{high}CD44^{high}$ $CD55^{high}$ cells are more able to form lumens than $CD24^{high}CD44^{high}$ cells within the lumen lines. It appears as though the addition of CD55 to the dual CD24/CD44 sort enriched the CSC population even more so and promotes the selection of cells that can both initiate clones

within three-dimensional growth matrices but can also differentiate into organized structures. Thus, on the basis of clonogenicity and lumen-formation, CD55 seems to be strongly linked to the CSC phenotype.

4.2.10 Adding CD55 to a CD24/CD44 dual flow cytometry sort further enhances proliferation

To test further the hypothesis that adding CD55 to a CD24/CD44 dual sort further enriches for cancer stem cells, sorted cells were evaluated for proliferation. Cell lines were sorted on the basis of CD24, CD44 and CD55 expression, separated into the most and least fluorescent 1% of the total population and grown in triplicate in 3,000-cell aliquots in 96-well plates for one week. Every two days, the level of cell proliferation was assessed by the MTS colorimetric assay. The results of the assay were plotted as growth curves (Figure 4.13), and the slopes of the curves were compared to ascertain the subpopulation with the greatest proliferative potential (Table 4.5).

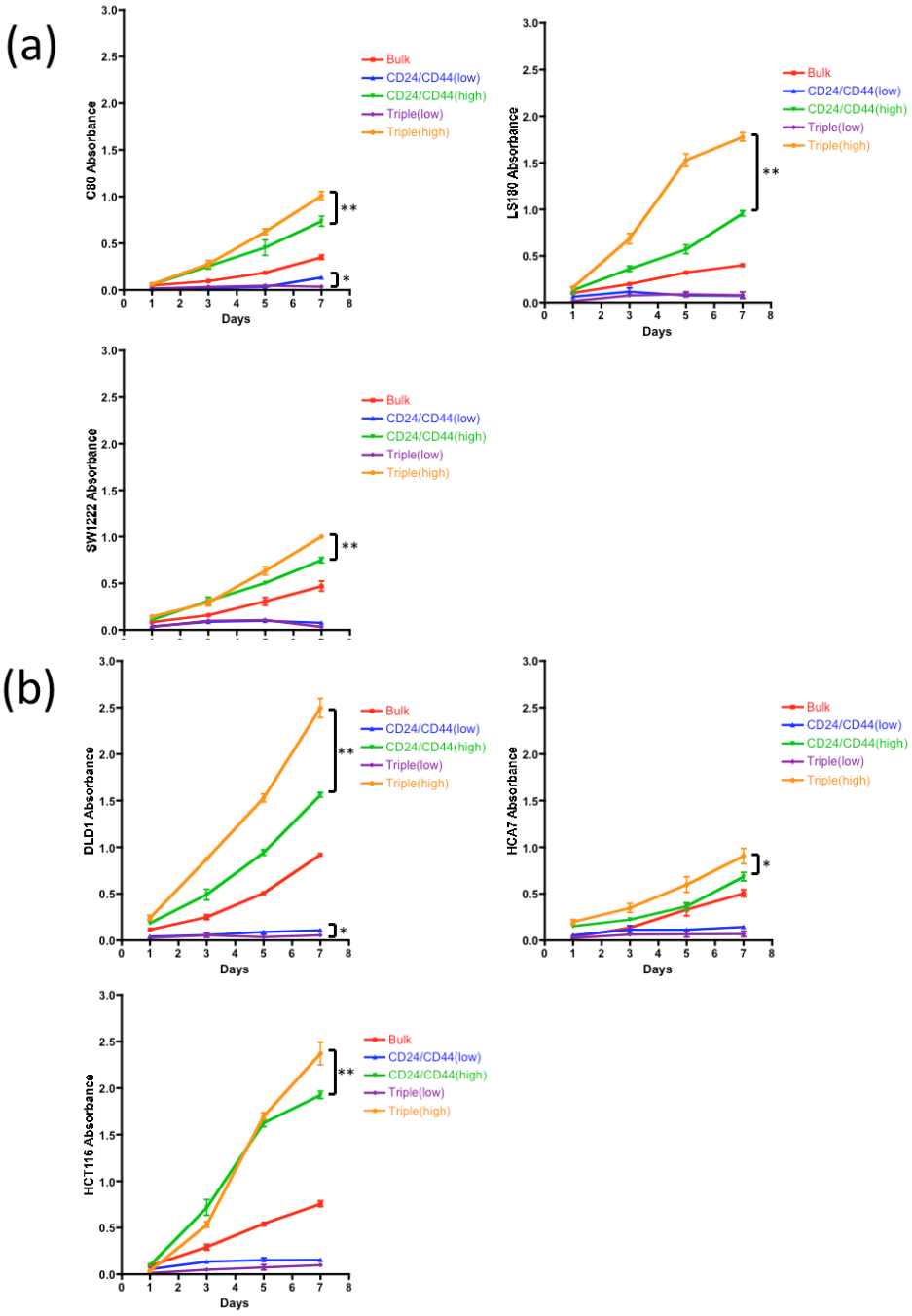


Figure 4.13 Growth curves corresponding to MTS-measured cell proliferation for sorted cell fractions from lumen (a) and dense (b) cell lines on the basis of CD24, CD44 and CD55 expression. Plotted points represent six data points (two experiments done in triplicate). For all six cell lines tested, the $CD24^{high}CD44^{high}CD55^{high}$ fraction was the most proliferative and grew the most rapidly over a seven-day test period, while the $CD24^{low}CD44^{low}CD55^{low}$ fraction was the least proliferative, as compared to their CD24/CD44 dual-sorted counterparts. The slopes of the sorted populations were compared to that of the unsorted, bulk, populations via Student's *t*-tests. * - 0.05 significance level, ** - 0.01 significance level, *** - 0.005 significance level.

Cell Line	$CD24^{low}$ $CD44^{low}$ Slope (units/day)	$CD24^{high}$ $CD44^{high}$ Slope (units/day)	$CD24^{low}$ $CD44^{low}$ $CD55^{low}$ Slope (units/day)	$CD24^{high}$ $CD44^{high}$ $CD55^{high}$ Slope (units/day)
C80	0.0190 ± 0.0078	0.1119 ± 0.0073	0.0037 ± 0.0021	0.1590 ± 0.0139
LS180	0.0007 ± 0.0067	0.1339 ± 0.0141	0.0099 ± 0.0056	0.2847 ± 0.0390
SW1222	0.0061 ± 0.0076	0.1054 ± 0.0040	0.0003 ± 0.0110	0.1454 ± 0.0183
DLD1	0.0120 ± 0.0011	0.2288 ± 0.025	0.0028 ± 0.0030	0.3706 ± 0.0284
HCA7	0.0136 ± 0.0039	0.0868 ± 0.0200	0.0065 ± 0.0031	0.1189 ± 0.0125
HCT116	0.0158 ± 0.0066	0.3198 ± 0.0404	0.0134 ± 0.0010	0.4086 ± 0.0435

Table 4.5 Summary of proliferation data from fluorescence-activated cell sorting on the basis of CD24, CD44 and CD55. Presented values represent six data points (two experiments done in triplicate). Among the lines, selection of the $CD24^{high}CD44^{high}CD55^{high}$ population yielded the highest growth rates, while selection of the $CD24^{low}CD44^{low}CD55^{low}$ population often reduced proliferation relative to its CD24/CD44 dual-sorted counterpart.

Among the three lumen lines, purification of the $CD24^{high}CD44^{high}CD55^{high}$ subpopulation enhanced proliferation by 50-100% over the $CD24^{high}CD44^{high}$ subpopulation, with LS180 showing the greatest increase and C80 the smallest. The $CD24^{high}CD44^{high}CD55^{high}$ subpopulations purified from the three dense lines grew more quickly than their lumen line counterparts (except for HCA7). The triple-sorted populations experienced only 33-50% increases in proliferation—HCA7 showed the smallest increase. Again, it appears as though the HCA7 population cannot be subdivided into groups of radically-different growth potentials as can be done for the other cell lines. Surprisingly, the dense lines showed smaller increases in proliferation than did the lumen lines; perhaps the dense lines are already extremely proliferative (as their bulk culture growth rates would confirm) and therefore, the addition of a third marker cannot enhance this high proliferation rate much more. Alternatively, there could still be some residual heterogeneity in the proliferative abilities of the $CD24^{high}CD44^{high}CD55^{high}$ cells

within the dense lines in which a significant minority of $CD24^{high}CD44^{high}CD55^{high}$ cells are still unable to grow quickly.

The $CD24^{low}CD44^{low}CD55^{low}$ cells were extremely slow-growing. Although in LS180, the population grew slightly faster than the $CD24^{low}CD44^{low}$ cells, the triple-sorted populations grew slightly more slowly than the $CD24^{low}CD44^{low}$ cells. In HCT116, the decrease was negligible, but in the other four, the growth rates dropped to 16-50% of their double-sorted counterparts.

The $CD24^{high}CD44^{high}CD55^{high}$ cells are able to grow at substantially higher rates than the $CD24^{high}CD44^{high}$ cells. Based upon this phenotype, CD55 appears to select for a fraction of the $CD24^{high}CD44^{high}$ population that is even more proliferative than the parental dual-sorted population. Clearly, CD55 enriches for CSCs, while the lack thereof dilutes the cancer stem cell population.

4.3 Discussion

The goal of most CSC studies is to identify a novel marker that can be associated with the cancer stem cell phenotype. To study CSC behavior, it is often necessary to isolate the putative cancer stem cell population in order to compare its phenotype with other populations or the bulk culture as a whole. While any marker can be used to accomplish this task, identifying a membrane protein with an extracellular domain is the best marker to choose, as it allows for fluorescently-linked or magnetic bead-linked antibodies to be used to separate subpopulations efficiently and quantitatively.

In the studies highlighted there, the CD55 membrane protein was identified as a potential

marker of cancer stemness, and a Fisher's exact test of the association between cell line morphology and *CD55* microarray mRNA expression confirmed this potential link between CD55 and cancer stemness. Quantitative reverse transcriptase PCR should be performed to confirm the link, as the microarray data of mRNA expression was not in complete concordance with the flow cytometry data of protein expression, particularly for the LS180 cell line, which was classified as a low *CD55* mRNA expresser yet expressed high levels of CD55 protein. While such a finding could possibly be explained by different levels of mRNA stability and translation among the cell lines, confirmatory PCR analysis would be a useful confirmation of the microarray expression data.

Six cell lines representing the two major colony morphologies (dense and lumen) were selected in order to ascertain the effects of CD55 purification on a diverse cell line panel. Since the CD55 expression profile, as revealed by flow cytometry, was quite varied, and for four of the cell lines, showed high levels of expression, the extremes of the CD55 distribution were purified and tested to assess the hypothesis that CD55 gradients matter more than a simple binary positive/negative determination. Cell sorting on the basis of CD55 alone revealed it to be a marker of clonogenicity, ability to differentiate and proliferation. Furthermore, CD55 was able to mark a subpopulation of self-renewing cells capable of differentiating. CD55 was capable of highlighting cells that exhibited an enlarged G_2/M cell cycle population. While Harper and colleagues (2010) found a proliferative CSC compartment within the G_2 phase cell population, not the G_2/M blended phase as found here, Harper's findings lend themselves to a further interrogation of the G_2/M phase cells for their potential phenotypic concordance to that of G_2 phase cells. Finally, intracellular probing with anti-CDX1 and anti-BMI1 antibodies showed that $CD55^{high}$ cells do indeed act similarly to stem cells, in that they overexpress BMI1 and

underexpress CDX1 compared to the bulk culture. The weight of *in vitro* evidence presented here has demonstrated that CD55 is a marker of cancer stemness.

Previous work within the CIL (Yeung, et al., 2010) found that the $CD24^{high}CD44^{high}$ subpopulation of colorectal cancer cells behaves as a CSC population. However, the putative CSC population was unable to generate 1,000 colonies when 1,000 cells were plated in Matrigel™, so there clearly was room for improvement in order to enrich the population for CSCs. In a bid to enhance the clone-forming abilities of the $CD24^{high}CD44^{high}$ population, the population was further purified with the addition of CD55. Selecting the top 1% of the distribution, thus purifying the $CD24^{high}CD44^{high}CD55^{high}$ population did enhance clonogenicity, ability to form lumens and proliferation all rose when the $CD24^{high}CD44^{high}$ was further purified. However, there is still much work to be done in order to approach a true CSC population where all 1,000 cells in an aliquot are able to give rise to 1,000 colonies. Some factors underlying this disconnect could be technical, such as the long length of the cell sorts that kept cells out of their proper growth medium or the cold temperature of the Matrigel™ that shocked cells normally accustomed to 37 ^{O}C growth temperatures. Thus, both biological (defining more faithful and rigorous markers of cancer stemness) and technical (reducing cell stress and challenges to cell viability) factors need to be optimized when designing experiments to isolate a pure CSC population.

CHAPTER 5

ROLES OF BMI1 AND CDX1 IN CANCER STEMNESS

5.1 Introduction

As previously discussed, separating colorectal cancer cell lines on the basis of CD55 expression revealed the membrane protein to be a potent marker of cancer stemness through a variety of cellular behavioral experiments—$CD55^{high}$ cells act like cancer stem cells (CSCs), and high CD55 expression enriches the $CD24^{high}CD44^{high}$ population even further for CSCs. Simultaneously, protein expression experiments on $CD55^{high}$ cells revealed that the transcription factor CDX1 was minimally-expressed and the chromatin regulatory factor BMI1 was strongly-expressed. CDX1 appeared to be negatively associated with the CSC population, while BMI1 appears to be positively associated. The results of the morphology microarray expression analysis further connected CDX1 to the differentiated lumen colony cell lines and BMI1 to the undifferentiated dense colony cell lines. BMI1 appears to be linked to cancer stemness, and CDX1 appears to be linked to cancer differentiation.

CDX1 encodes the CDX1 transcription factor that promotes the transcription of genes related to cellular differentiation, including those related to MUC2 and E-cadherin, and represses the transcription of genes related to cell proliferation, including those related to β-catenin and cyclin D2 (Guo, et al., 2004). Its functions have been best-characterized in the large bowel, where CDX1 promotes the proper patterning and differentiation of colonic crypts—luminal colonic epithelial cells express high levels of CDX1, while inferior colonic crypt base cells express low levels. Since the stem cell compartment within the colon is thought to reside at the crypt base, CDX1 is positively associated with epithelial differentiation (Chan, et al., 2009). Furthermore, analyses of colorectal cancer cell lines indicated that methylation of the *CDX1* promoter significant dampens *CDX1* transcription in a significant minority of cell lines, thus leading to a loss of protein expression (Wong, et al., 2004).

BMI1 encodes the BMI1 chromatin regulatory protein that associates with the RING1B, PH1 and CBX4 proteins to form the PRC1 aggregate (Gil, et al., 2005). PRC1 modifies gene transcription on the chromatin level to silence target genes associated with cellular differentiation (Gil, et al., 2005), thus maintaining a stem-like phenotype. These effects have been observed in a diverse array of tissues, but within the colon and rectum, high expression of BMI1 has been associated with tumorigenesis (Li, et al., 2010)—within a panel of biopsied colorectal tissues, a higher proportion of tumors stained positive for BMI1 than did normal epithelium.

Previous experiments with CD55 were made possible by virtue of the molecule's presence on the cell membrane, allowing for antibody-based flow cytometry analyses that directly probe for cellular expression without need for cell permeabilization, lysis or transfection. However, CDX1 and BMI1 are intracellular proteins, largely precluding flow cytometry and FACS as useful tools. Other methodologies, such as transgene transfection, could potentially be used to ascertain the roles of the two proteins in promoting or dampening CSC populations or their associated phenotypes.

5.2 Results

5.2.1 *BMI1* and *CDX1* mRNA expression levels vary among colorectal cancer cell lines

The results of the Bonferroni-corrected Student's *t*-test comparing mean BMI1 and CDX1 mRNA expression levels of dense cell lines to lumen cell lines indicated a statistically-significant increase in expression associated with the formation of dense colonies. Previous RT-PCRs performed by Wong and colleagues (2004) indicated that a significant

minority, seven, of a panel of thirty-seven colorectal cancer cell lines expressed little *CDX1* mRNA, so it would be likely that *CDX1* mRNA expression from the microarray panel could be partitioned into two groups of cell lines. To determine if this is the case, BMI1 and CDX1 mRNA expression was plotted against cell lines in rank order (Figure 5.1) to assess any discontinuities in the distribution.

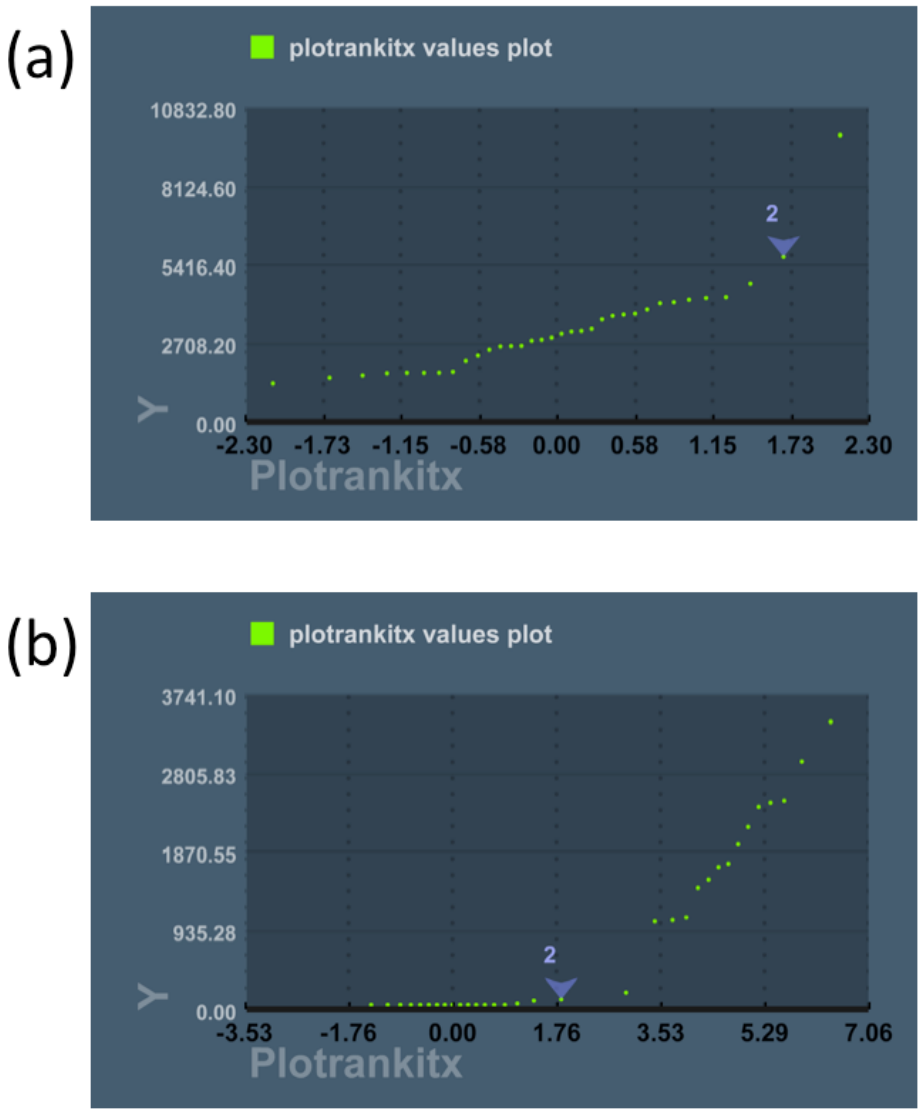


Figure 5.1 Plots of *BMI1* (a) and *CDX1* (b) microarray expression data versus rankits. For *BMI1*, the data range was a relatively narrow one order of magnitude, and the algorithm partitioned the panel into two groups at an expression level of 7740, with the second group consisting of just one cell line, SKCO1. For CDX1, the data range was a wide three orders of magnitude, and the algorithm partitioned the panel into two groups at an expression level of 82. The *p*-value cutoff was 0.001, the F-test cutoff was 0.01 and the coefficient of variation was 0.3.

Applying the rankit regression algorithm developed by the CIL to the *CDX1* expression plot allowed for the separation of the cell line panel. At a significance level of 0.001, one discontinuity in the line plot was observed between the HCA7 and SKCO1 cell lines at an expression level of 82 fluorescence units (midway between the expression level of HCA7, 57, and of SKCO1, 106). The nine lumen cell lines tended to appear on the lower end of the distribution, while the seven dense cell lines tended to appear on the higher end.

To test the association between *CDX1* expression level and cell line colony morphology, a Fisher's exact test was performed on the partition of dense and lumen cell lines into high and low *CDX1* expression classifications (Figure 5.2). One lumen line and five dense lines were low for *CDX1*, while two dense lines and eight lumen lines were high for *CDX1*. The resulting 2 x 2 table indicated that *CDX1* expression was positively associated with the formation of lumen lines at a *p*-value of 0.0245. The statistical significance of the Fisher's exact test corroborates the statistical significance of a Student's *t*-test comparing the mean expression values of the dense and lumen lines (*p*-value of 0.0006).

p-value = 0.0245	Low *CDX1* Expression	High *CDX1* Expression
Dense Cell Lines	**5**	**2**
Lumen Cell Lines	**1**	**8**

Figure 5.2 2 x 2 table of cell lines classified by morphology and by level of *CDX1* expression. At a *p*-value of 0.0245, there is an association between low *CDX1* expression and the dense colony morphology and between high *CDX1* expression and the lumen colony morphology.

A different story exists for *BMI1* expression. At a significance level of 0.001, one discontinuity in the line plot was observed between the Vaco5 and SKCO1 cell lines at an expression level of 7740 fluorescence units (midway between the expression level of Vaco5, 5631, and of SKCO1, 9848). All nine lumen lines and seven dense lines were below the partition point, so a Fisher's exact test on the basis of the algorithm results would not be appropriate. However, ranking the cell lines by *BMI1* expression and partitioning the set into terciles reveals a deeper pattern. Among the lowest tercile lie seven lumen lines (CCK81, C84, HRA19, LS513, SW1222 and LIM1863) and two dense lines (HCA7 and WIDR). Among the highest tercile lie no lumen lines and four dense lines (LS411, HCT116, HCT15 and DLD1). The resulting Fisher's exact test (Figure 5.3) on the resulting classification indicated that *BMI1* expression was positively associated with the formation of dense lines at a *p*-value of 0.0210. The statistical significance of the Fisher's exact test corroborates the statistical significance of a Student's *t*-test comparing the mean expression values of the dense and lumen lines (*p*-value of 0.0155).

5.2.2 *BMI1* and *CDX1* cDNA transgene expression can persist in cell lines

To assess the effects of *BMI1* and *CDX1* on the behavior of colorectal cancer cell lines, two plasmid vector constructs, pBI-CMV3-BMI1 (Figure 5.4 (a)) and pBI-CMV3-CDX1 (Figure 5.4 (b)), were synthesized that drove expression of an inserted cDNA sequence (here, sequences corresponding to *BMI1* and *CDX1*). The utilized vector possessed a bidirectional CMV promoter with enhancer to drive expression of the transgene and the reporter ZsGreen1 green fluorescent protein to allow for *in situ* tracking of plasmid transfection and expression. The ColE1 origin of replication within the plasmid allowed for propagation in prokaryotes, while the SV40 origin of replication within the plasmid allowed for propagation in eukaryotes.

p-value = 0.0210	Very Low *BMI1* Expression	Very High *BMI1* Expression
Dense Cell Lines	2	4
Lumen Cell Lines	7	0

Figure 5.3 2 x 2 table of cell lines classified by morphology and by level of *BMI1* expression. At a *p*-value of 0.0210, there is an association between very low *BMI1* expression and the lumen colony morphology and between very high *BMI* expression and the dense colony morphology.

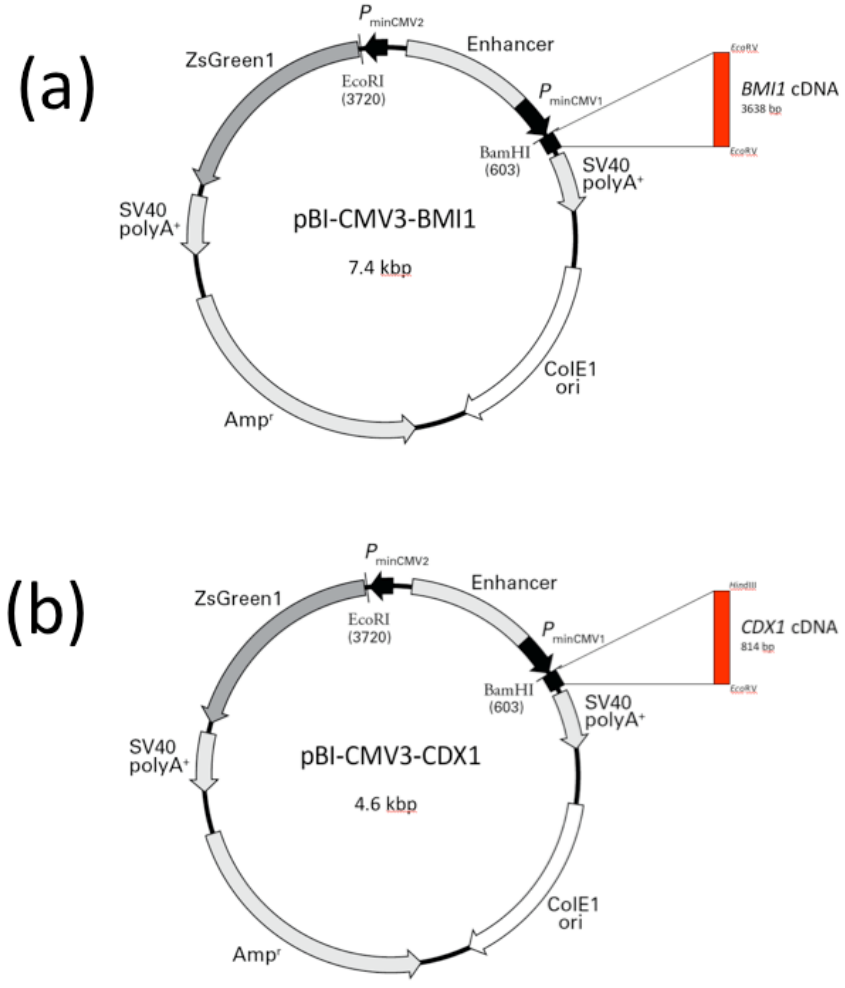


Figure 5.4 Constructed plasmid vectors (Clontech, 2009) incorporating the *BMI1* (a) and *CDX1* (b) cDNAs that are constitutively expressed alongside a reporter green fluorescent protein.

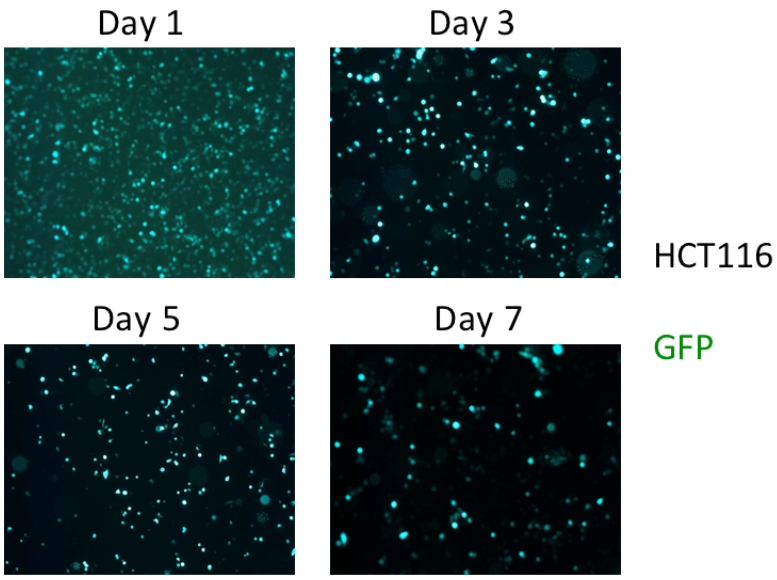


Figure 5.5 Example photographs (10x magnification, 488-nm fluorescence excitation) of HCT116 cells transfected with the pBI-CMV3-CDX1 experimental vector. Even after seven days, a significant minority of cells still expressed the vector.

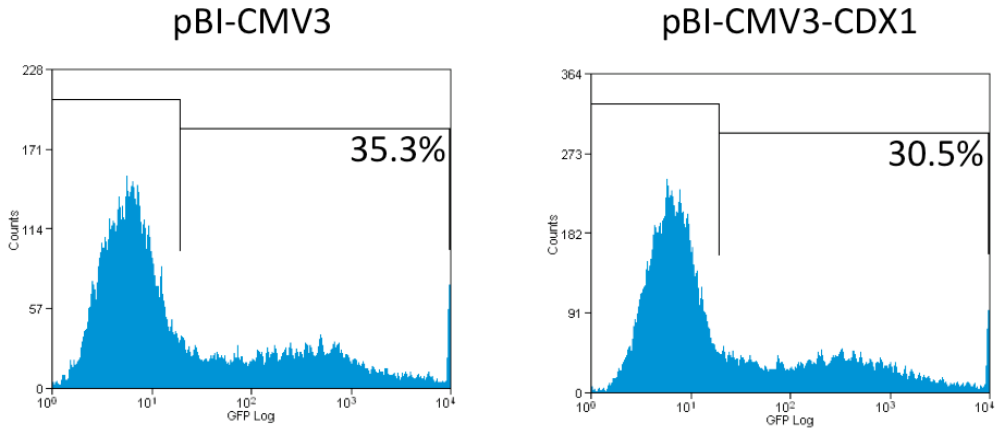


Figure 5.6 Flow cytometry histograms of HCT116 cells transfected with pBI-CMV3 control or pBI-CMV3-CDX1 experimental vectors. The most fluorescent cells, corresponding to the right gates, were the cells that had successfully taken up the transfected plasmid and thus were purified via FACS.

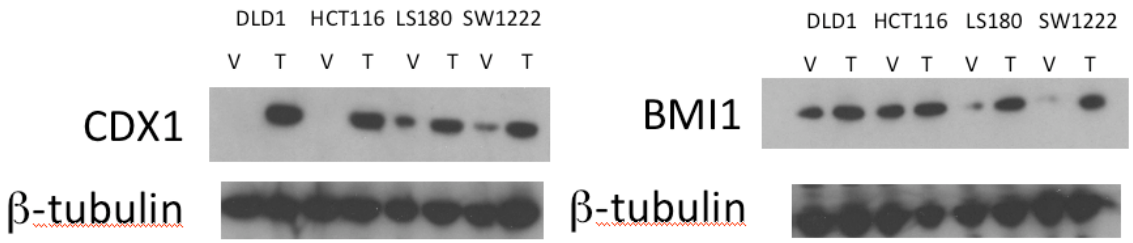


Figure 5.7 Western blots showing induction of CDX1 expression (left) and of BMI1 expression (right) within transfected cells. DLD1 and HCT116 expressed high levels of BMI1 endogenously (V), but transfection (T) with the experimental pBI-CMV3-BMI1 did induce an increase in BMI1 expression.

The lumen cell lines LS180 and SW1222 as well as the dense cell lines DLD1 and HCT116 were selected for transfection experiments. Cells were harvested and separated into aliquots of 10^5 cells each for plating in 24-well plates. Cells were subsequently chemically transfected with 1 μg control or experimental vector DNA using the Lipofectamine™ LTX with PLUS™ lipofection system to assess the effects of forced expression of *BMI1* and *CDX1* on cell behavior, such as clone formation or proliferation.

To evaluate transfection efficiency, cultures were photographed 24 hrs post-transfection with an inverted fluorescence microscope to assess the proportion of cells expressing the reporter ZsGreen1 green fluorescent protein. All four cell lines exhibited high transfection efficiencies. To evaluate transfection persistence, cultures were subsequently photographed over one week to assess the prevalence of cells expressing the fluorescent reporter. An example photographic time course for the HCT116 cell line is presented in Figure 5.5.

To evaluate the effectiveness of the transfection procedure to induce expression of BMI1 or CDX1 protein, transfected cultures were sorted on the basis of ZsGreen1 expression via FACS (an example for the HCT116 cell line is presented in Figure 5.6), and the ZsGreen1$^+$ cells (which successfully took up the transfected plasmid) were purified and lysed. Resulting cellular lysates were electrophoresed, transferred to a nitrocellulose membrane and probed for expression of CDX1 and BMI1 with monoclonal antibodies via Western blotting (Figure 5.7). A monoclonal antibody against β-tubulin was used as a loading control to establish that differences in protein expression were not attributable to variations in the amount of total protein loaded. Western blotting revealed that the cells transfected with vectors containing a cDNA insert expressed higher levels of BMI1 or CDX1 than did cells transfected with an empty, control vector. Among the two dense

lines DLD1 and HCT116, induction of BMI1 expression was not as strong as it was among the two lumen lines LS180 and SW1222, but an induction was observed nonetheless. CDX1 induction was extremely strong across all four cell lines. Thus, populations of cells expressing heightened levels of BMI1 and CDX1 proteins were successfully created and purified, allowing for downstream analyses of cell behavior as a function of these enforced higher protein expression levels.

5.2.3 *BMI1*-transfected and *CDX1*-transfected cell populations exhibit different clonogenicity propensities

To assess the clonogenic abilities of the transfected cell populations expressing higher levels of *BMI1* or *CDX1* than normal, transfected cells were sorted on the basis of ZsGreen1 expression, and the ZsGreen1^{+} population was purified. Transfected cells were grown in triplicate in 1,000-cell aliquots in Matrigel™ in 96-well plates for two weeks to assess their clonogenic abilities in three-dimensions. Although transgene expression was shown for one week, the initiation of clones and/or of differentiated luminal structures begins within hours or days of cell plating—the major events that influence clonogenicity would have occurred while the *BMI1* or *CDX1* cDNA was expressed highly.

After two weeks of culture, wells were photographed (Figure 5.8) and the number and type of colonies within each well was assessed (Figure 5.9 and Table 5.1). Examples of distinctive colonies identified within the pBI-CMV3-CDX1-transfected dense lines are presented in Figure 5.10.

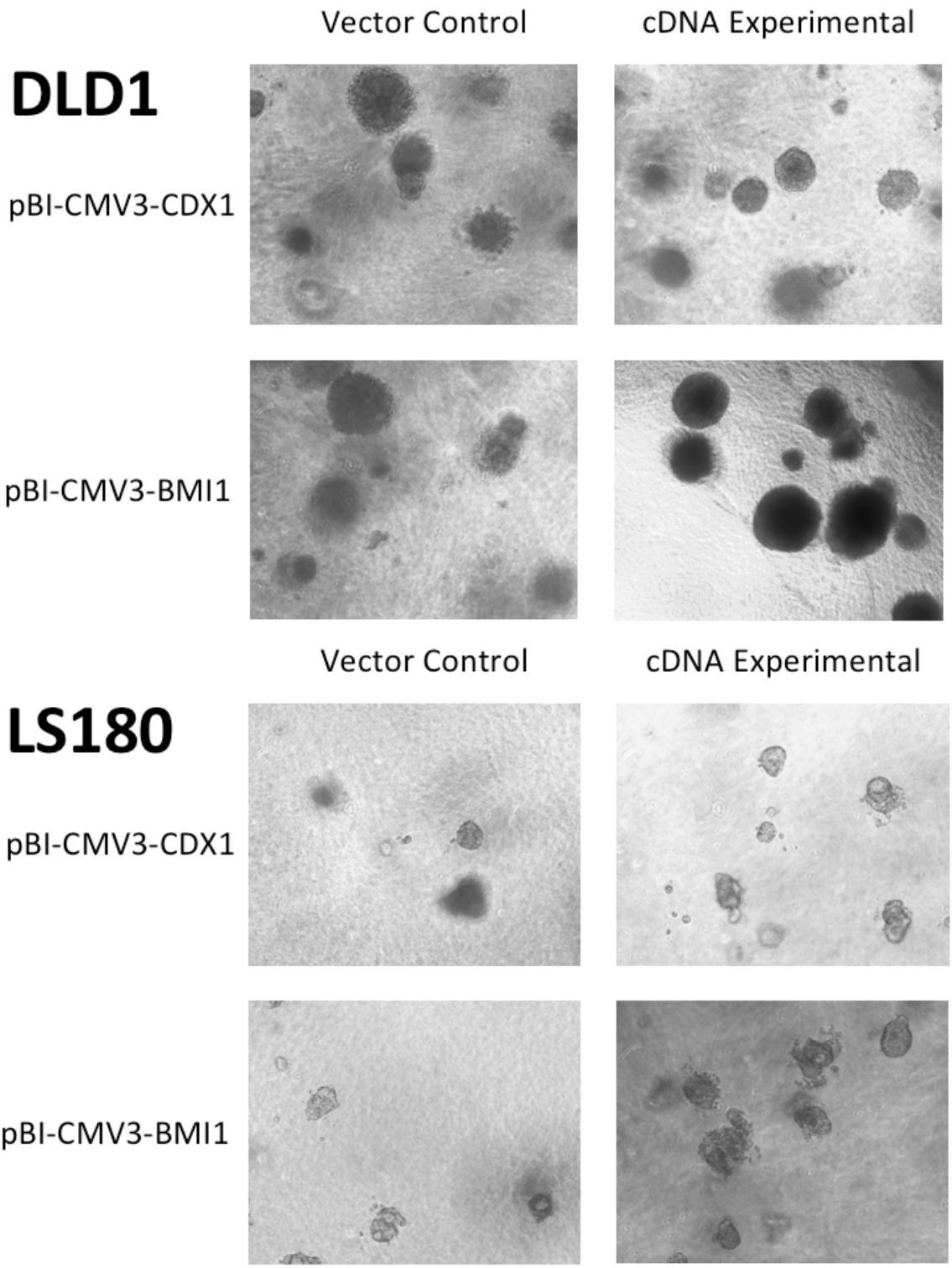


Figure 5.8 Example photographs (10x magnification under phase contrast) of 1,000 sorted cells from the DLD1 (a) and LS180 (b) cell lines in wells of a 96-well plate. *CDX1* appeared to dampen clonogenicity and enhance the formation of lumens. *BMI1* appeared to enhance clonogenicity, while within lumen lines, appeared to dampen the formation of lumens.

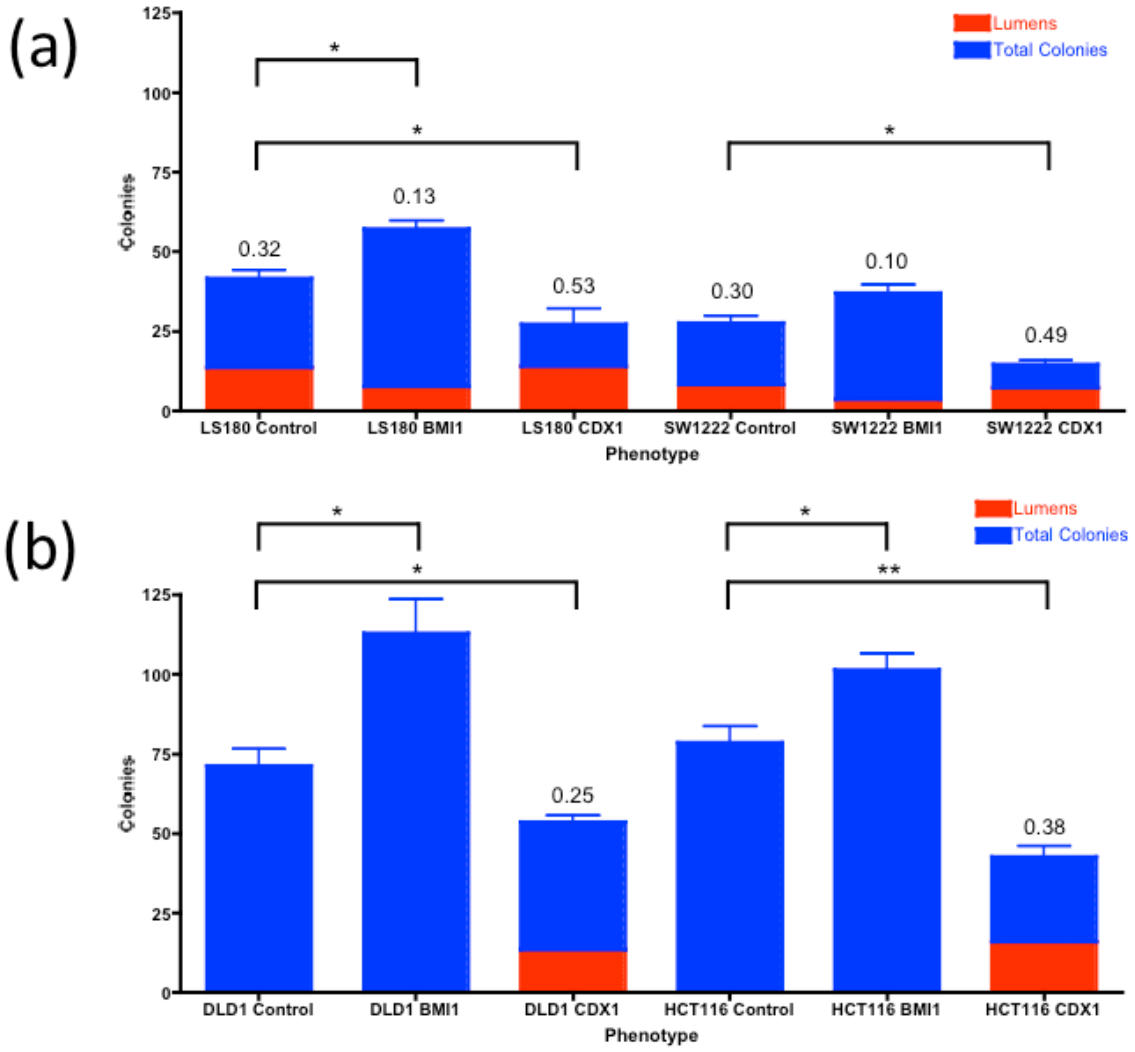


Figure 5.9 Bar graphs showing the number and morphology of colonies (from Table 5.1) generated by transfected lumen cell line cells (a) and dense cell line cells (b). Bars represent six data points (two experiments done in triplicate). Cell lines were purified on the basis of ZsGreen1 reporter expression, divided into 1,000-cell aliquots and grown in Matrigel™ for two weeks. Among the lumen lines, transfection of *BMI1* enhances clonogenicity and dampens lumen formation, while transfection of *CDX1* dampens clonogenicity and enhances lumen formation. Among the dense lines, transfection of *BMI1* enhances clonogenicity, while transfection of *CDX1* dampens clonogenicity and, interestingly, stimulates the formation of nascent lumen-containing colonies where none existed before. The mean proportions of lumens were compared to the unsorted, bulk population via Student's *t*-tests; all were significant at a 0.01 significance level. The mean numbers of colonies were compared to the unsorted, bulk population via Student's *t*-tests (brackets shown on graph). * - 0.05 significance level, ** - 0.01 significance level.

Cell Line	pBI-CMV3		pBI-CMV3-BMI1		pBI-CMV3-CDX1	
	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens
LS180	42 ± 5.5	0.32 ± 0.04	57 ± 4.1	0.13 ± 0.02	27 ± 4.5	0.53 ± 0.10
SW1222	28 ± 3.7	0.30 ± 0.02	37 ± 3.5	0.10 ± 0.02	15 ± 2.9	0.49 ± 0.04
DLD1	71 ± 5.2	0.00	113 ± 10.6	0.00	54 ± 5.2	0.25 ± 0.03
HCT116	79 ± 5.0	0.00	101 ± 5.2	0.00	43 ± 5.0	0.38 ± 0.04

Table 5.1 Summary of clonogenicity data from cell transfections. Presented values represent six data points (two experiments done in triplicate). Among the lumen lines, transfection of *BMI1* enhances clonogenicity and dampens lumen formation, while transfection of *CDX1* dampens clonogenicity and enhances lumen formation. Among the dense lines, transfection of *BMI1* enhances clonogenicity, while transfection of *CDX1* dampens clonogenicity and stimulates the formation of nascent lumen-containing colonies.

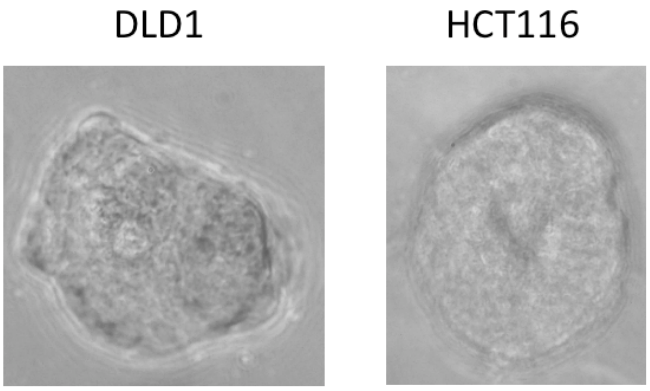


Figure 5.10 Example photographs (10x magnification under phase contrast) of colonies from *CDX1*-transfected DLD1 and HCT116 cells. DLD1 and HCT116 were unable to differentiate *in vitro* to give rise to organized, polarized lumen-like structures, but the introduction of *CDX1* clearly induced this ability.

The lumen cell lines showed moderate increases in clonogenicity (although not at a statistically-significant level in SW1222) and, interestingly, decreases in the proportion of colonies that form lumens upon constitutive induction of *BMI1* expression. Both LS180 and SW1222 saw their clone-forming abilities rise by more than 30%. Furthermore, *BMI1* seemed to have interfered with the abilities of LS180 and SW1222 to differentiate and

self-organize into lumen-like structures—transfection of pBI-CMV3-BMI1 forced a decline of 60% in lumen formation among the two cell lines. Among the dense lines, increases were also observed in clonogenicity of the *BMI1*-transfected cells of 25% or more, slightly lower than those experienced by the lumen lines, possibly as a result of the high endogenous expression of *BMI1* within DLD1 and HCT116.

Upon transfection with *CDX1* cDNA, the lumen lines showed statistically significant decreases in clonogenicity and increases in the proportion of lumen-containing colonies. SW1222 saw its clone-forming ability drop by more than 50%, while LS180 saw a smaller drop of 35%. Furthermore, *CDX1* seemed to have enhanced the ability of colonies to differentiate, as evidenced by a rise of more than 50% within both lines in lumen-forming ability. Among the dense lines, decreases in clonogenicity were also noticed on a similar level to those experienced by the lumen lines (25-50%). What is possibly the most interesting observation emerging from these experiments is the development of lumen-forming colonies among the dense cell lines transfected with *CDX1*. Formerly undifferentiated, unpolarized colonies now, upon induction of *CDX1* expression, begin to differentiate, as was recently observed with the HCT116 cell line when stably transfected with *CDX1* cDNA (Yeung, et al., 2010).

Increased *BMI1* expression can be associated with both an increase in CSC proliferation (as evidenced by the increases in clonogenicity) and with a decrease in cellular sensitivity to serum-mediated differentiation signals (as evidenced by the decreases in lumen formation). Increased *CDX1* expression can be associated with both a decrease in CSC proliferation (as evidenced by the decreases in clonogenicity) and with an increase in cellular sensitivity to serum-mediated differentiation signals (as evidenced by the increases in lumen formation). Thus, on the basis of clonogenicity, *BMI1* appears to be

linked to the cancer stem cell phenotype while *CDX1* appears to be linked to the cancer differentiated cell phenotype.

5.2.4 *BMI1*-transfected and *CDX1*-transfected cell populations exhibit different proliferative propensities

The proliferative aspect of the CSC phenotype can also be evaluated within the context of *BMI1* and *CDX1* transfection to ascertain the possibility that forced increases in BMI1 or CDX1 protein expression can drive or halt the growth of a tumor. To assess such abilities *in vitro*, cells were transfected with pBI-CMV3, pBI-CMV3-BMI1 or pBI-CMV3-CDX1 and purified via FACS using the reporter ZsGreen1 green fluorescent protein. Isolated, transfected cells were grown in triplicate in 3,000-cell aliquots in 96-well plates for one week. Every two days, the level of cell proliferation was assessed by the MTS colorimetric assay. The results of the assay were plotted as growth curves (Figure 5.11), and the slopes of the curves were compared to ascertain the subpopulation with the greatest proliferative potential (Table 5.2).

Cell Line	pBI-CMV3 Culture Slope (units/day)	pBI-CMV3-BMI1 Slope (units/day)	pBI-CMV3-CDX1 Slope (units/day)
LS180	0.0945 ± 0.0040	0.1623 ± 0.0180	0.0510 ± 0.0017
SW1222	0.0920 ± 0.0044	0.1573 ± 0.0155	0.0485 ± 0.0054
DLD1	0.1389 ± 0.0390	0.2164 ± 0.0284	0.0704 ± 0.0199
HCT116	0.1366 ± 0.0127	0.1720 ± 0.0143	0.1002 ± 0.0148

Table 5.2 Summary of proliferation data from vector transfections. Presented values represent six data points (two experiments done in triplicate). Among all cell lines, the *BMI1*-transfected cultures were the most proliferative and grew the most rapidly over a seven-day test period, while the *CDX1*-transfected cultures were the least proliferative.

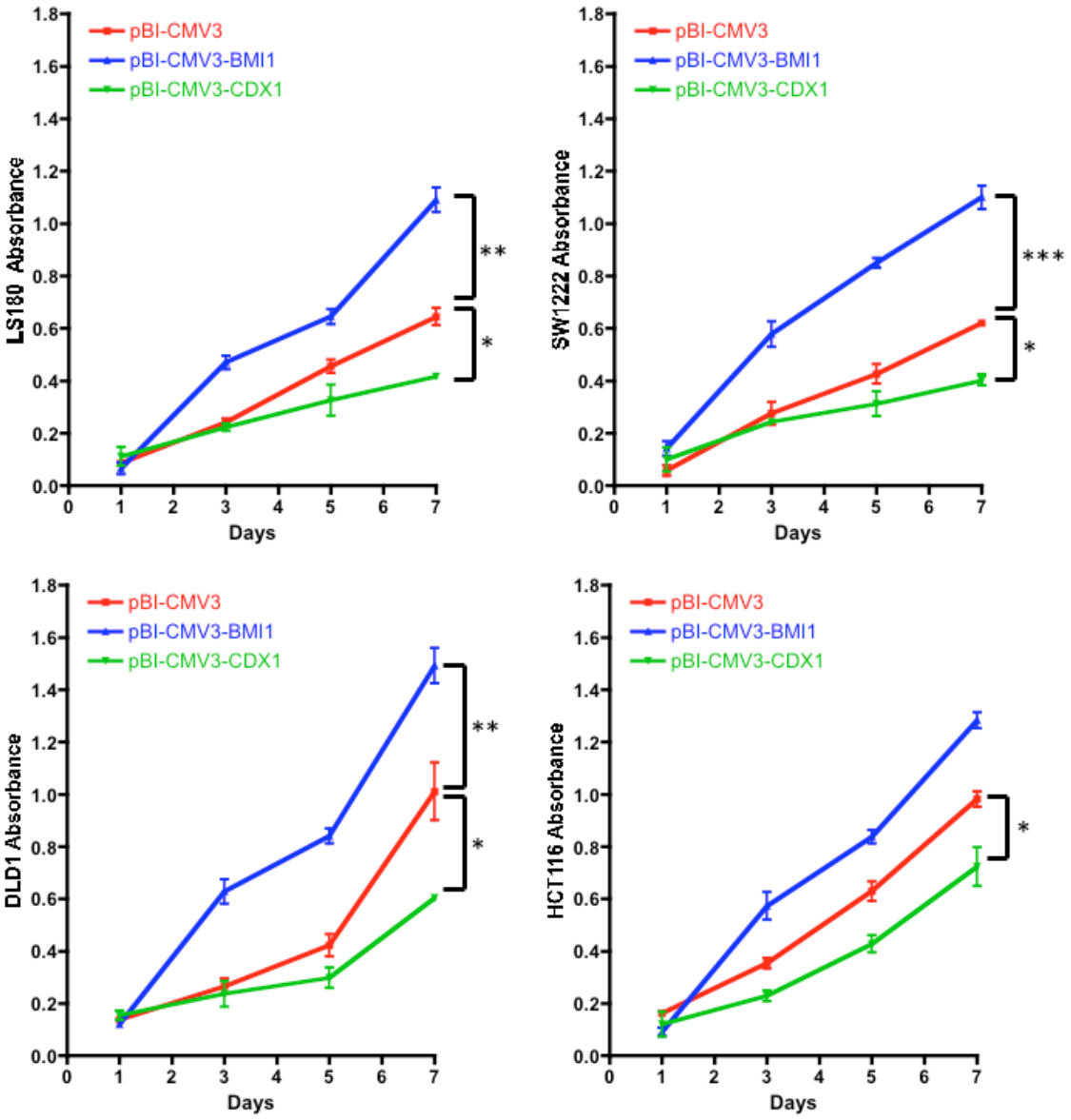


Figure 5.11 Growth curves corresponding to MTS-measured cell proliferation for transfected cells. Plotted points represent six data points (two experiments done in triplicate). For all four cell lines tested, the BMI1-transfected cultures were the most proliferative and grew the most rapidly over a seven-day test period, while the CDX1-transfected cultures were the least proliferative. The slopes of the sorted populations were compared to that of the unsorted, bulk, populations via Student's *t*-tests. * - 0.05 significance level, ** - 0.01 significance level, *** - 0.005 significant level.

Transfection of pBI-CMV3-BMI1 enhanced proliferation at statistically significant levels for both lumen lines, while transfection of pBI-CMV-CDX1 dampened proliferation at statistically significant levels. For both dense lines, transfection of pBI-CMV3-BMI1 enhanced proliferation significantly for DLD1, but not for HCT116, while both lines exhibits significant decreases in proliferation upon transfection with pBI-CMV3-CDX1.

A potential cause for this differential effect could be related to the overall control processes within the cells. Within the lumen lines, their abilities to differentiate in Matrigel™ imply that differentiation control pathways are active. Induced *BMI1* expression would be able to silence the transcription of genes associated with differentiation, so it is likely that induced *BMI1* could have potent effects in differentiating cells. However, in the dense lines, where stemness pathways are likely active, an increase in *BMI1* expression would not change the overall trend in the stemness control program—thus, an increase in *BMI1* could yield a small-to-moderate increase in proliferation, which was observed.

Increased *BMI1* expression can be associated with an increase in CSC proliferation (as evidenced by the increases in proliferation), while increased *CDX1* expression can be associated with a decrease in CSC proliferation (as evidenced by the decreases in proliferation).

5.2.5 CDX1^{high} cells can be labeled *in situ*

While the previous experiments showed that increased *CDX1* expression is associated with reduced colony formation and increased propensity to respond to differentiation cues in three-dimensional culture as well as with reduced cell proliferation, the system by which these conclusions were derived was somewhat artificial. *CDX1* expression was

induced by means of a plasmid transfection—the cells originally did not express such high levels of *CDX1*, and such high levels are not normally found in colonic epithelial cells. To assess the role of endogenous *CDX1* expression in cancer stem cell biology, a plasmid vector construct, pZsGreen1-1-CDX1_Prom (Figure 5.12) was synthesized that drove expression of a reporter green fluorescent protein, ZsGreen1, via an 1154-base upstream fragment of the CDX1 promoter. The pUC origin of replication within the plasmid allowed for propagation in prokaryotes, while the SV40 origin of replication within the plasmid allowed for propagation in eukaryotes. The utilized vector drove expression of ZsGreen1 when the inserted *CDX1* promoter was active, allowing for *in situ* tracking of promoter (and thus transcriptional activity of *CDX1*) activation. ClustalW sequence alignment of the upstream regions of CDX1 from human and mouse (Figure 5.13 (a)) revealed that a high degree of similarity existed between the two regions within 1,000 bases of the transcriptional start site (Figure 5.13 (b)). Thus, a promoter fragment roughly corresponding to the first 1,000 upstream bases was synthesized and ligated into the reporter vector. Although a potential enhancer region was detected 3,000 bases upstream by virtue of a plateau in alignment scores, the first 1,000 bases showed a much higher similarity.

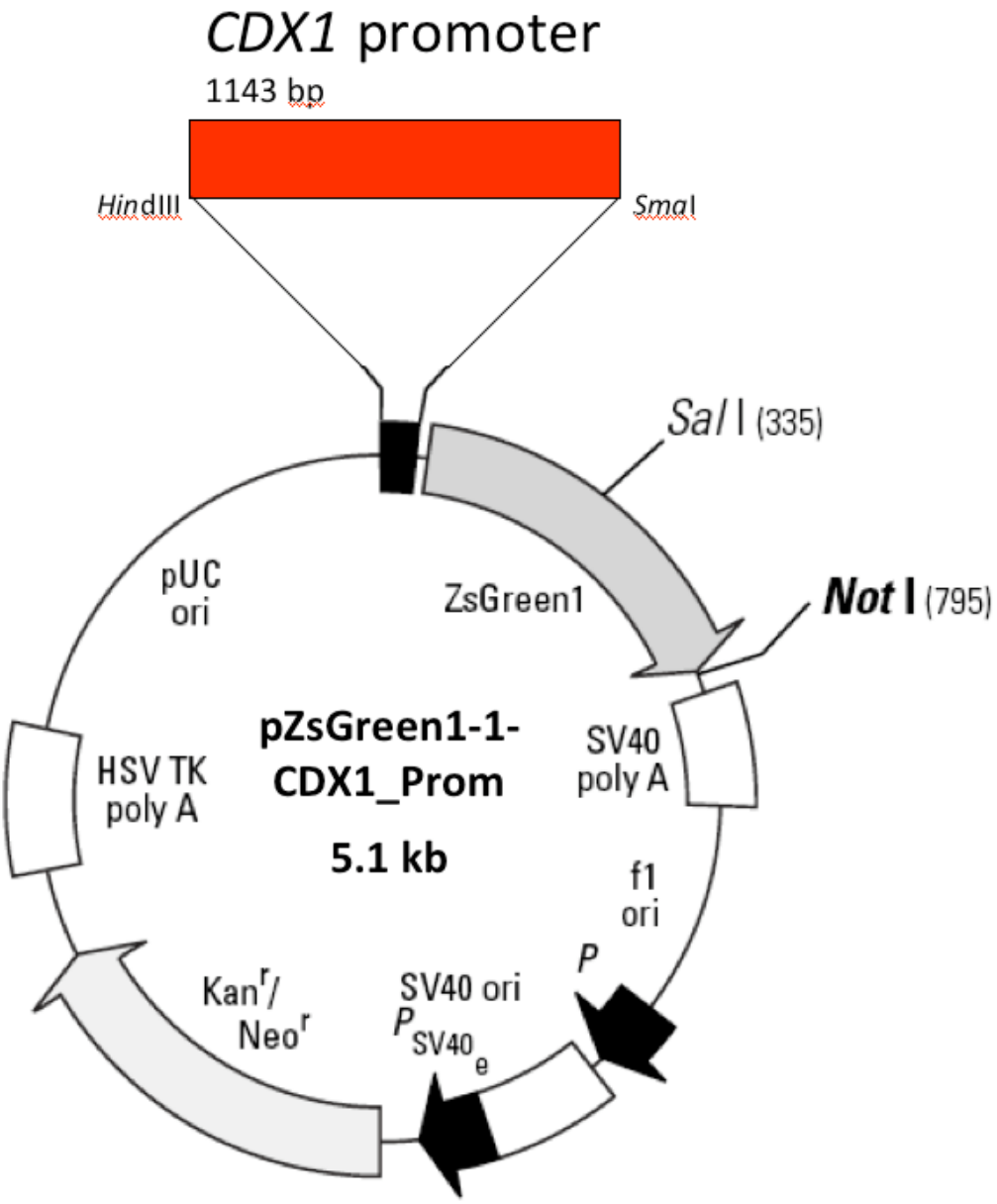


Figure 5.12 Constructed plasmid vector (Clontech, 2004) incorporating the *CDX1* promoter fragment that drives expression of a downstream reporter green fluorescent protein.

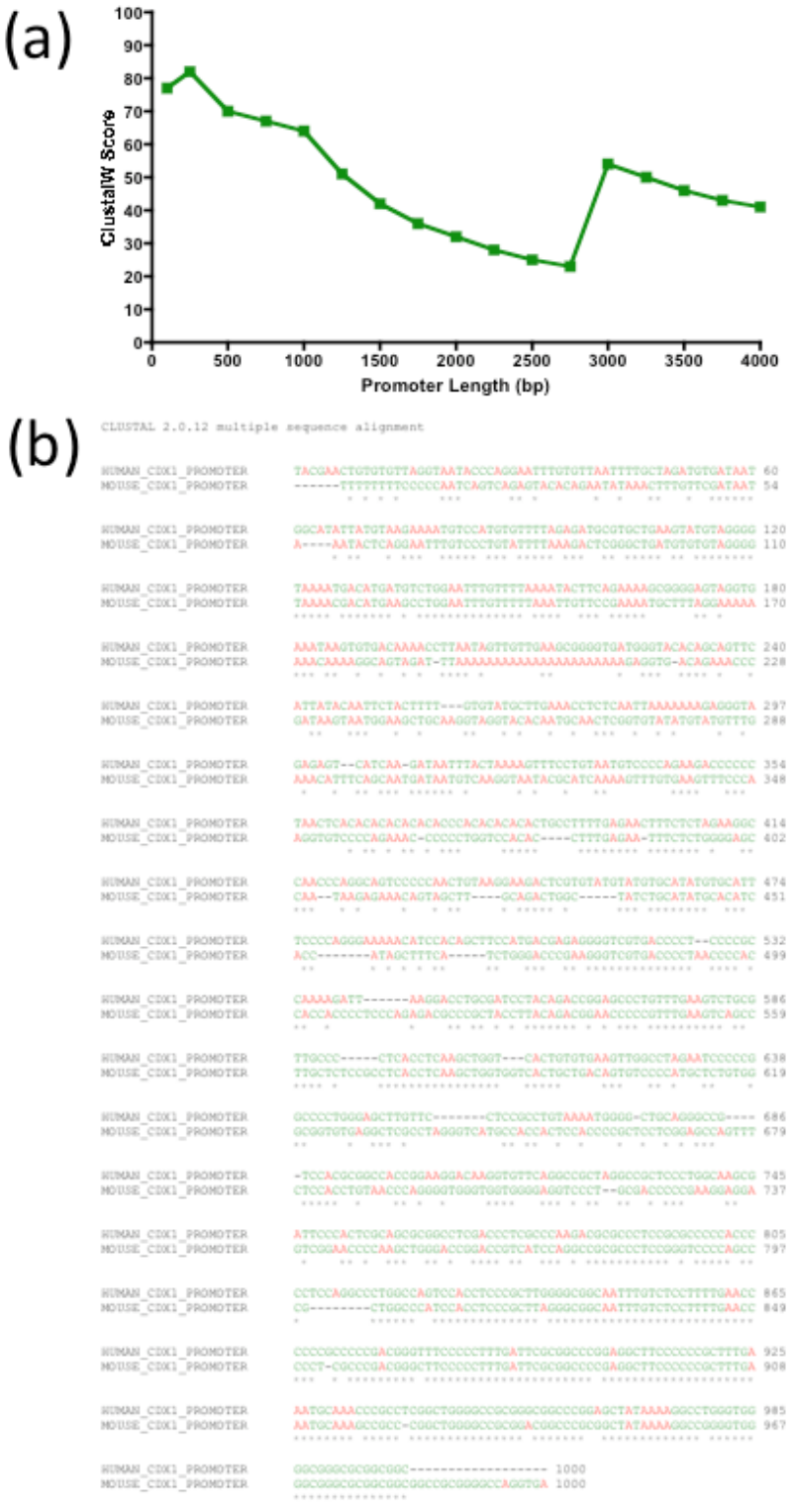


Figure 5.13 ClustalW results from an alignment of the *CDX1* upstream promoter regions from the human and mouse genomes. A graph of the alignment scores (a) shows a general trend downward as the promoter region is expanded upstream, with plateaus 1,000 bases and 3,000 bases upstream. Alignment of the first 1,000 upstream bases (b) revealed significant conservation of sequence between human and mouse, implying that the region is likely to be a main site of *CDX1* transcriptional control.

The lumen cell line SW1222 was selected for transfection experiments, as its transcriptional expression of *CDX1*, as revealed by the microarray expression analysis, was high. Cells were harvested and separated into aliquots of 10^5 cells each for plating in 24-well plates. Cells were subsequently chemically transfected with 1 μg control (p3gE) or experimental (pZsGreen1-1-CDX1_Prom) vector DNA using the Lipofectamine™ LTX with PLUS™ lipofection system. The control p3gE contained a 3.4 kb upstream promoter fragment of the pan-intestinal epithelial gene *TACSTD1* (encoding the EpCAM protein) that drove expression of a green fluorescent protein (GFP) reporter within the vector pEGFP-1 (McLaughlin, et al., 2004)—if transfected, the vector promoter should be ubiquitously activated, making p3gE a good transfection control. McLaughlin and colleagues validated the control vector within the colorectal cancer cell line SW498. Transfection efficiency, per p3gE plasmid expression, was approximately 70%.

Transfected cells were counterstained with anti-CD55 antibody and analyzed via flow cytometry to ascertain the proportion and distribution of CD55 expression within the pZsGreen1$^+$ (either EpCAM$^+$ or CDX1$^+$) transfected cells (Figure 5.14).

The SW1222 cell line experiment that tested the pZsGreen1-1-CDX1_Prom vector construct within the context of CD55 flow cytometric expression revealed that CDX1$^+$ cells could be labeled *in situ* via promoter-reporter construct transfection. Since expression of the intracellular CDX1 protein could then be detected by flow cytometry via GFP, transfected cells were counterstained with CD55 antibody to ascertain the level of CD55 expression among CDX1$^+$ cells.

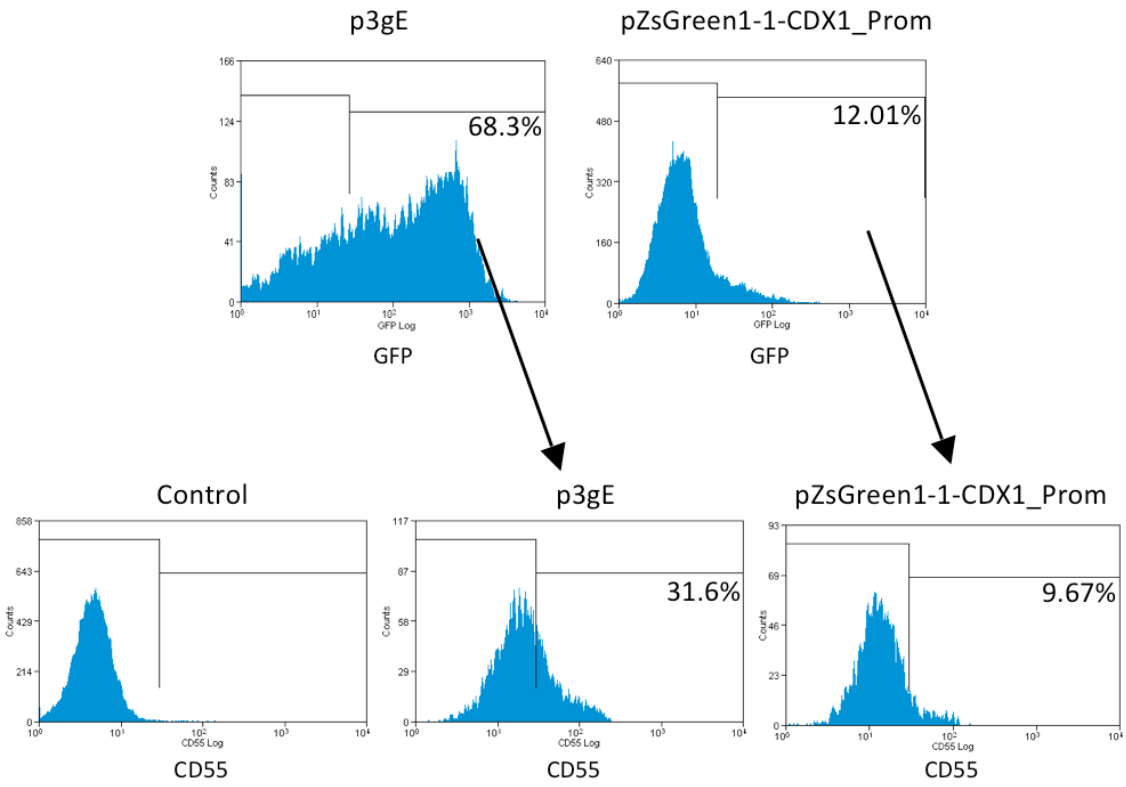


Figure 5.14 Transfection and subsequent flow cytometric analysis of SW1222 cells with a control *TACSTD1* (EpCAM) promoter plasmid p3gE and an experimental *CDX1* promoter plasmid pZsGreen1-1-CDX1_Prom. The *CDX1* promoter-reporter construct was able to label a proportion of the population that was expressing CDX1. Furthermore, counterstaining with CD55 revealed that the $CDX1^+$ cells were less positive for CD55 than were $EpCAM^+$ cells.

The control cells, a representation of the bulk culture, were 31.6% positive for CD55, while the experimental cells, a representation of the $CDX1^+$ differentiated subpopulation within the bulk culture, were only 9.67% positive. Although this 9.67% figure is similar to the positive percentage previously quoted for SW1222, the cells here were not fixed prior to flow cytometry analysis (fixation inhibits detection of the GFP reporter), so it is possible that the lack of fixation intensified the fluorescent signal from the CD55 antibody—hence, the control cells were a high 31.6% positive. The results here indicated that, at least for SW1222, $CDX1^+$ cells expressed far less CD55 than the bulk culture, implying a link between CDX1 expression and a loss of the $CD55^+$ CSC population.

5.3 Discussion

While CSC studies often focus on identifying surface markers useful for isolation of positive and negative populations, the question of the regulatory control processes underpinning CSCs and their differentiated counterparts is often left unanswered. The markers used to isolate CSCs could, in fact, have little functional relevance in stem cell or differentiated cell regulator processes. Indeed, a recent study by Horst and colleagues (2009) found that while the putative CSC marker CD133 was associated with a poorer prognosis in colorectal cancer, the ablation of CD133 expression via RNA interference had little biological or behavioral effect, calling into question the role of CD133 in maintaining a CSC phenotype. Previously discussed microarray results linking *CDX1* expression to differentiated, lumen-forming colonies and *BMI1* expression to undifferentiated, dense colonies, as well as results linking BMI1 expression as well as the lack of CDX1 expression to the $CD55^{high}$ subpopulation, suggest that BMI1 and CDX1 are regulators of colorectal columnar cell stemness and differentiation. Thus, they were examined more closely.

In the studies highlighted herein, two cDNA vectors were constructed that drove expression of BMI1 or CDX1 protein alongside a green fluorescent protein reporter. Vectors were transfected into two lumen and two dense lines, and the cells positive for the reporter were isolated and assessed for CSC properties, such as clonogenicity, lumen formation and proliferation. Three-dimensional growth assays of the transfected cells revealed *BMI1* to be an important promoter of cancer stemness and an important inhibitor of cell sensitivity to differentiation cues, in that enforced *BMI1* expression enhanced clone formation and dampened the formation of differentiated lumen structures, while the effects of *BMI1* were less pronounced on the already undifferentiated dense lines.

Simultaneously, CDX1 proved to be important in inhibiting cancer stemness and promoting cell sensitivity to differentiation cues among the lumen lines—clonogenicity was reduced, while lumen formation was increased. Between the two dense cell lines, induced CDX1 expression led to a more moderate decrease in clonogenicity, but it triggered the differentiation of the colonies, leading to the formation of polarized, organized lumen-containing colonies. The effects of *BMI1* and *CDX1* on proliferation followed the same pattern—cells grew less when transfected with a *CDX1* cDNA vector and grew more when transfected with a *BMI1* cDNA vector. Although further experiments, such as siRNA knockdowns or *in vivo* mouse xenografts, could be performed, an interesting experiment to perform first would be an assessment of BMI1 expression in *CDX1* cDNA-transfected cells and of CDX1 expression in *BMI1* cDNA-transfected cells. Since BMI1 is known to regulate homeobox genes (Gil, et al., 2005), of which *CDX1* is one, enforced *BMI1* expression would likely decrease CDX1 expression.

To test further the role of endogenous *CDX1* in cancer stemness, a green fluorescent protein reporter vector was constructed incorporating a 1154- base region of the CDX1 promoter that drove expression of the reporter. Transfecting the vector into the SW1222 cell line allowed for the identification of cells positive for CDX1 via flow cytometry, a novel solution that avoids problems related to the cell permeabilization required to use flow cytometry for intracellular proteins. Counterstaining transfected cells with anti-CD55 antibody revealed that $CDX1^{+}$ cells expressed less CD55, establishing a link between differentiation and a dampening of CD55 expression.

In conclusion, BMI1 has been shown to be linked to an expansion of the CSC population (through clonogenicity and proliferation assessments) and a decrease in differentiation potential, while CDX1 has been shown to be linked to a contraction of the CSC population (through clonogenicity, proliferation and CD55-staining assessments) and an increase in differentiation potential.

CHAPTER 6

ROLES OF MYOFIBROBLASTS IN CANCER STEMNESS

6.1 Introduction

Although it is the bedrock of tissue culture, cell monoculture often presents a different growth environment from the environment that cells experience *in vivo*. For colorectal cancer, a variety of cells exist within or near to a solid tumor, including the actual tumor (epithelial) cells, myofibroblasts, immune cells, smooth muscle cells and endothelial cells, all of which may influence the initiation, progression, invasion and architecture of a tumor (Marx, 2008). However, *in vitro* studies of cancer cells, especially cancer stem cells (CSCs), often occur in isolation, such as in the absence of tumor companion cells, such as myofibroblasts, which could influence cancer cell behavior.

Within the large bowel, the myofibroblast population likely plays a role in influencing colorectal cancer progression. Characterized by their hybrid smooth muscle/fibroblast morphology and immunohistochemistry (they stain positive for the smooth muscle marker α-smooth muscle actin and the fibroblast marker vimentin), myofibroblasts form a continuous sheath surrounding the colonic crypts and luminal epithelial tissues and lie along the extracellular matrix (ECM) in which the epithelial cells are embedded. Although not directly in contact with the epithelial cells, myofibroblasts regulate epithelial cell proliferation and crypt patterning through the synthesis of ECM molecules (including collagens I-IV and laminin) and the secretion of growth factors and cytokines, including the eponymous basic fibroblast growth factor, insulin-like growth factor I and interleukin-6. Indeed, the effects of these secreted factors are often mediated by ECM molecules through the binding of the factors to the ECM (Powell, et al., 1999a). Within the context of cancer, it is likely that these myofibroblasts are critical in tumor progression. For example, Denys and colleagues (2008) found that cultured HCT-8/E11 colorectal cancer cells became more invasive when grown superior to growth-factor-

activated primary colonic myofibroblasts. Indeed, within the context of CSC biology, extant evidence suggests a role for myofibroblasts. In colon cancer, CCD-18Co myofibroblasts appear to promote the stemness phenotype in that isolated CSC visually appear to remain undifferentiated in myofibroblast-conditioned medium (Vermeulen, et al., 2010).

Although few other studies have specifically analyzed colonic myofibroblasts (especially normal colonic myofibroblasts), experiments with cancer-associated breast fibroblasts (Huang, et al., 2010) and cancer-associated lung fibroblasts (Wang, et al., 2009) found a supportive role. In breast cancer, co-culture of the MCF7 cell line with fibroblasts increased the proportion of $CD44^{+}CD24^{-}$ cells (the putative CSC compartment) eight-fold and increased *in vivo* mouse xenograft tumorigenicity, while in lung cancer, co-culture of the PC-9 and HCC827 cell lines with fibroblasts induces cancer chemoresistance to the EGFR tyrosine kinase inhibitor gefitinib.

Previous experiments with BMI1, CD55 and CDX1 have focused on the responses of the colorectal cancer cell line panel in monoculture. By conducting co-culture experiments of intestinal myofibroblasts with epithelial cells, particularly within a three-dimensional growth matrix, the *in vitro* growth conditions begin to approach those of an *in vivo* tumor, allowing for a more detailed analysis of the mechanisms controlling CSC populations.

6.2 Results

6.2.1 The cell line CCD-18Co is myofibroblast-derived

Since primary myofibroblasts obtained from colonic biopsies are difficult to source in large quantities for extensive *in vitro* experiments and could potentially differ from

patient to patient, the CCD-18Co cell line was purchased from ATCC—the cells were derived from primary human colonic myofibroblast tissue and were not immortalized (Valentich, et al., 1997). To confirm their myofibroblast nature, CCD-18Co cells were grown on cover slips and stained immunofluorescently (Figure 6.1) with the anti-myofibroblast monoclonal antibody PR2D3 that has been shown to stain myofibroblast cells uniquely (Richman, et al., 1987). Furthermore, CCD-18Co cells were harvested, stained with an anti-EpCAM monoclonal antibody and analyzed via flow cytometry to confirm the lack of any colonic epithelial cell contamination in the cell line.

The experiments revealed that CCD-18Co cells are indeed PR2D3-positive myofibroblasts that are not contaminated with epithelial cells. The EpCAM-negative nature of the CCD-18Co cells allows for easy flow cytometric analyses of myofibroblast/epithelial co-culture experiments—myofibroblast populations can be excluded such that data collection can focus on phenotypic changes in the epithelial cell compartment.

6.2.2 Cell lines co-cultured with CCD-18Co cells exhibit enhanced clonogenicity propensities

To assess the effects of myofibroblast co-culture on the clonogenic abilities of four cell lines (the LS180 and SW1222 lumen lines and the DLD1 and HCT116 dense lines), epithelial cells were harvested, divided into 1,000-cell aliquots and combined with 0, 500, 1,000 or 2,000 myofibroblasts (producing myofibroblast : epithelial cell plating ratios of 0, 0.5, 1 and 2). All cell co-cultures were grown in calcium-containing DMEM, which would tend to favour the adhesion of cells to each other through E-cadherin interactions. Cell combinations were grown in triplicate in Matrigel™ in 96-well plates for two weeks to assess their clonogenic abilities in three-dimensions. Hypothetically, a more CSC-like

population should be able to generate more colonies. After two weeks of culture, wells were photographed (Figure 6.3) and the number and type of colonies within each well was assessed (Table 6.1 and Figure 6.4).

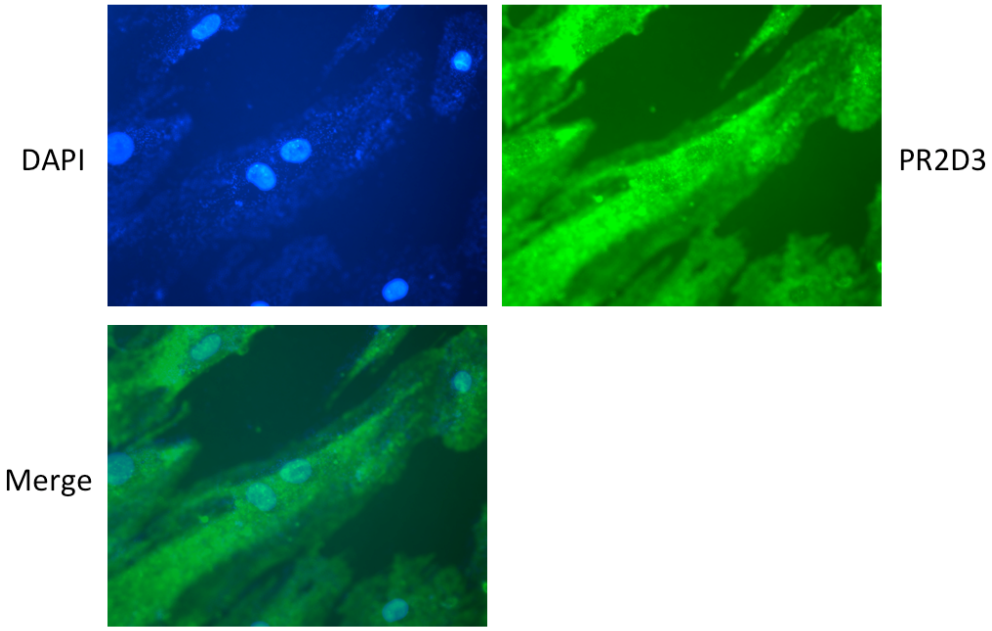


Figure 6.1 Immunofluorescence images (40x magnification) of CCD-18Co cells grown on cover slips. Cells stained positive for PR2D3, indicating that they are, in fact, myofibroblasts.

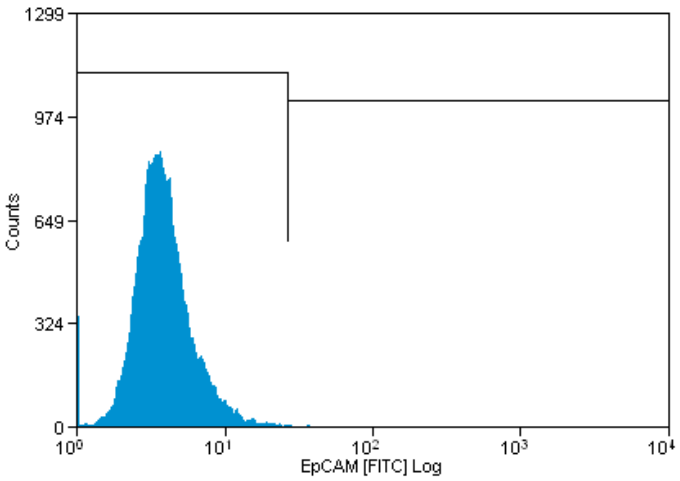


Figure 6.2 Flow cytometry histogram of CCD-18Co cells stained with anti-EpCAM antibody. CCD-18Co cells do not express the colonic epithelial marker EpCAM, indicating that no epithelial cells have contaminated the cell line.

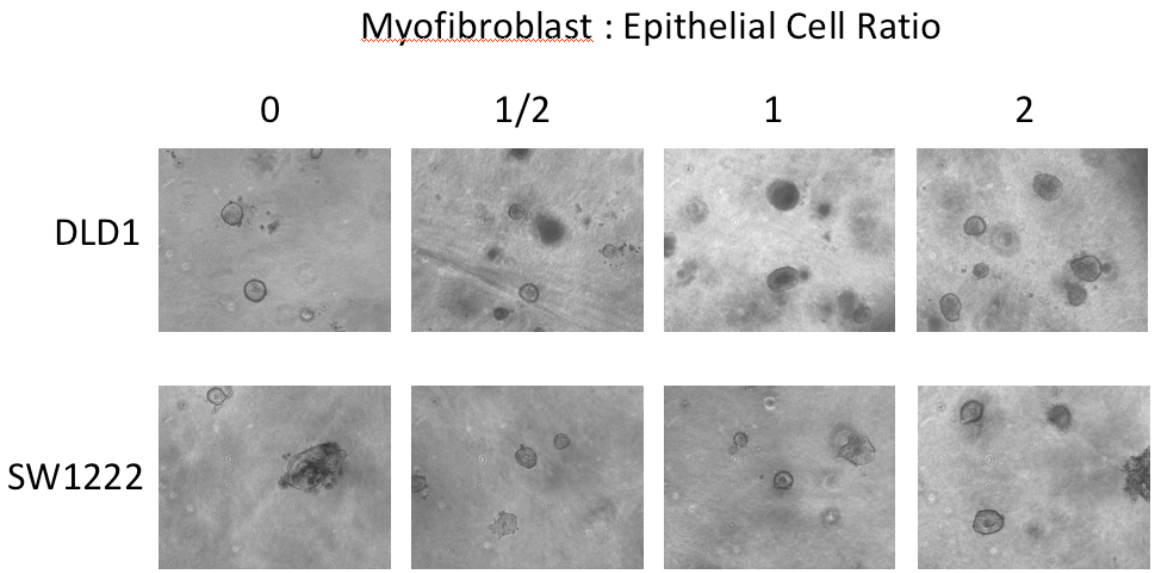


Figure 6.3 Example photographs (10x magnification under phase contrast) of 1,000 DLD1 or SW1222 cells co-cultured with varying ratios of myofibroblasts (0, 500, 1,000 or 2,000 cells) in 96-well plates. Increased numbers of myofibroblasts appear to enhance clonogenicity and, between the two lumen lines, appear to enhance the formation of lumens.

Cell Line	Myofibroblast : Epithelial Cell Ratio							
	0		0.5		1		2	
	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens	Colonies	Proportion Lumens
LS180	42 ± 2.8	0.29 ± 0.02	61 ± 6.4	0.48 ± 0.01	75 ± 3.8	0.63 ± 0.08	89 ± 2.6	0.59 ± 0.04
SW1222	30 ± 3.2	0.38 ± 0.03	44 ± 5.2	0.44 ± 0.02	62 ± 3.6	0.59 ± 0.05	75 ± 5.5	0.63 ± 0.09
DLD1	68 ± 4.2	0.00	103 ± 9.9	0.00	125 ± 8.4	0.00	151 ± 6.4	0.00
HCT116	78 ± 4.4	0.00	106 ± 4.1	0.00	122 ± 11	0.00	139 ± 2.6	0.00

Table 6.1 Summary of clonogenicity data from co-cultures. Presented values represent six data points (two experiments done in triplicate). Between the lumen lines, co-culture of myofibroblasts with epithelial cancer cells enhances clonogenicity and lumen formation—the enhancement appears to be dose-dependent. Between the dense lines, co-culture also enhances clonogenicity, but the effects of the myofibroblasts do not appear to be striking as the number of cells was increased.

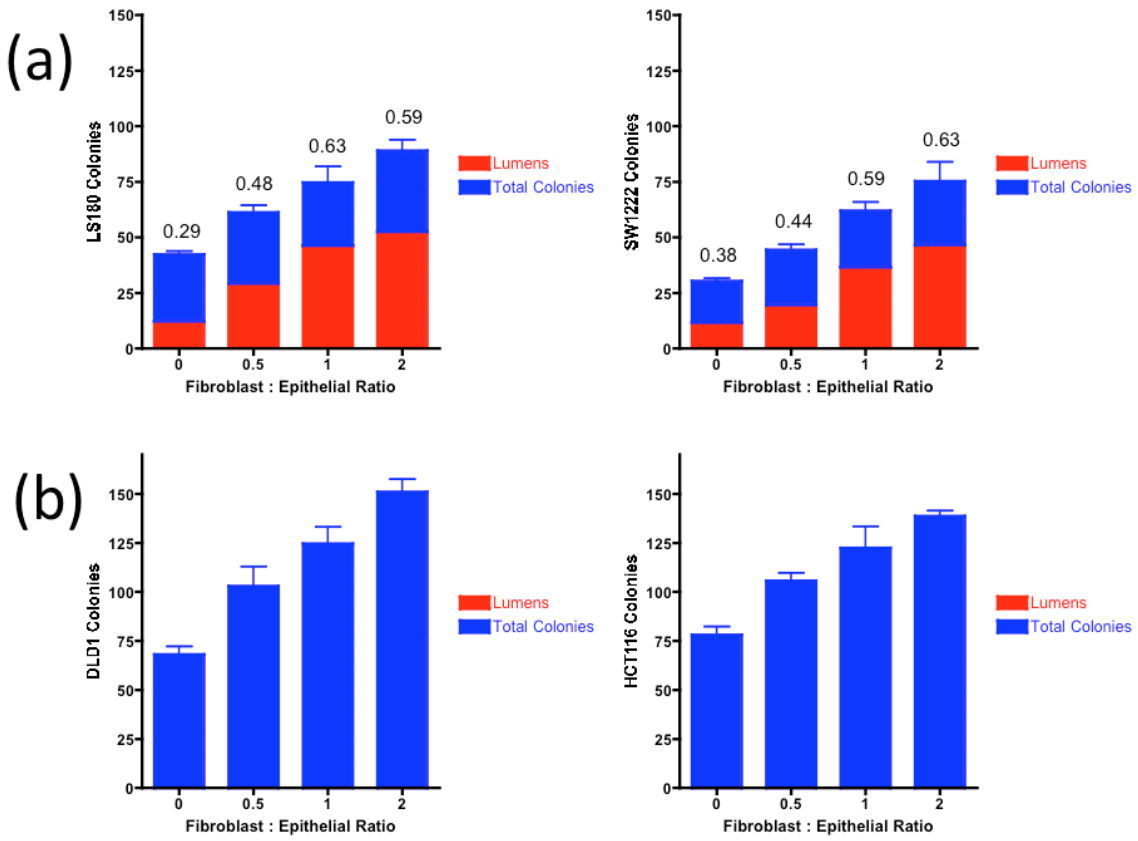


Figure 6.4 Bar graphs showing the number and morphology of colonies generated by myofibroblast co-cultured with lumen (a) and dense (b) cell lines. Bars represent six data points (two experiments done in triplicate). Cell lines were co-cultured with varying ratios of myofibroblasts (0, 500, 1,000 or 2,000 cells) in Matrigel™ in 96-well plates. Increased numbers of myofibroblasts appear to enhance clonogenicity and, within the two lumen lines, appear to enhance the formation of lumens.

Co-culture of the lumen lines with myofibroblasts led to a doubling in clone-forming abilities as the number of CCD-18Co cells increased to twice the number of epithelial cells. Furthermore, the proportions of colonies possessing lumens also doubled for both lumen lines as the number of myofibroblasts co-cultured increased. Between the two dense lines, co-culture also led to a near-doubling in clone-formation as the number of CCD-18Co cells increased to twice the number of epithelial cells, but no induction of lumen formation was observed.

The number of colonies formed, and for the lumen lines the proportions of colonies possessing lumens, appeared to show a dose dependence to the number of myofibroblasts. To tease out this relationship, regressions of clonogenicities and lumen proportions (Figure 6.5 and Table 6.2) were plotted and *t*-tests for slope computed to identify the dependence, if any, between the number of myofibroblasts and clonogenicity/lumen formation.

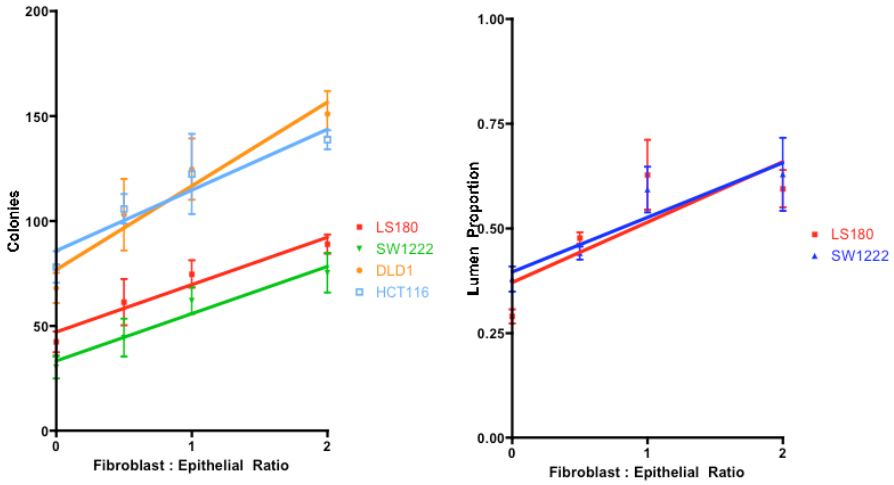


Figure 6.5 Regression plots of clonogenicity (left) and proportion of colonies with lumens (right) versus the myofibroblast/epithelial cell ratio. At a significance level of 0.005, all four plots of clonogenicity versus ratio deviated from zero, while at a significance level of 0.01, both plots of lumen proportion versus ratio deviated from zero.

Cell Line	Clonogenicity Regression Slope (clones/ratio unit)	Lumen Proportion Regression Slope (proportion/ratio unit)
LS180	22.59 ± 3.08	0.1440 ± 0.0446
SW1222	22.55 ± 3.15	0.1303 ± 0.0366
DLD1	39.92 ± 5.51	
HCT116	28.99 ± 4.78	

Table 6.2 Summary of slopes from regressions of clonogenicity versus myofibroblast/epithelial cell ratio and lumen proportion versus myofibroblast/epithelial cell ratio. At a significance level of 0.005, all four plots of clonogenicity versus ratio deviated from zero, while at a significance level of 0.01, both plots of lumen proportion versus ratio deviated from zero.

The regressions indicated that all four cell lines were sensitive to myofibroblast-mediated induction of clonogenicity at a significance level of 0.005, with the two dense lines approximately 33% more sensitive than the two lumen lines. Furthermore, the two lumen lines were sensitive to induction of lumen-forming colonies at a 0.01 significance level.

Myofibroblast co-culture enhanced proliferation of CSCs, as evidenced by the observed increases in clonogenicity across all four cell lines, and enhanced sensitivity to differentiation, as evidenced by the observed increases in lumen proportion within the two lumen lines. Co-culture did not induce the formation of lumens within the dense lines, likely a result of dense line cells unable to respond to exogenous differentiation cues from the environment.

6.2.3 Cell lines co-cultured with CCD-18Co cells exhibit enhanced proliferation propensities

In vitro CSC proliferation can also be evaluated within the context of myofibroblast co-culture. Epithelial cells were harvested, divided into 3,000-cell aliquots and combined with 3,000 myofibroblasts. Cell combinations were grown in triplicate in 96-well plates for one week. Every two days, cell proliferation was assessed by the MTS colorimetric assay. Assay results were plotted as growth curves (Figure 6.6), and the slopes of the curves were compared to ascertain the populations with the greatest proliferative potential (Table 6.3). Because the addition of the myofibroblasts to the well would automatically increase measured proliferation, CCD-18Co cells were also grown in isolation—the resulting MTS absorbance data from the CCD-18Co cells alone was subtracted from the co-culture absorbance data to arrive at figures that were reflective of the epithelial cell proliferation, a mathematical transform that assumes no pro-proliferative effect on myofibroblast growth when co-cultured with colorectal cancer epithelial cells.

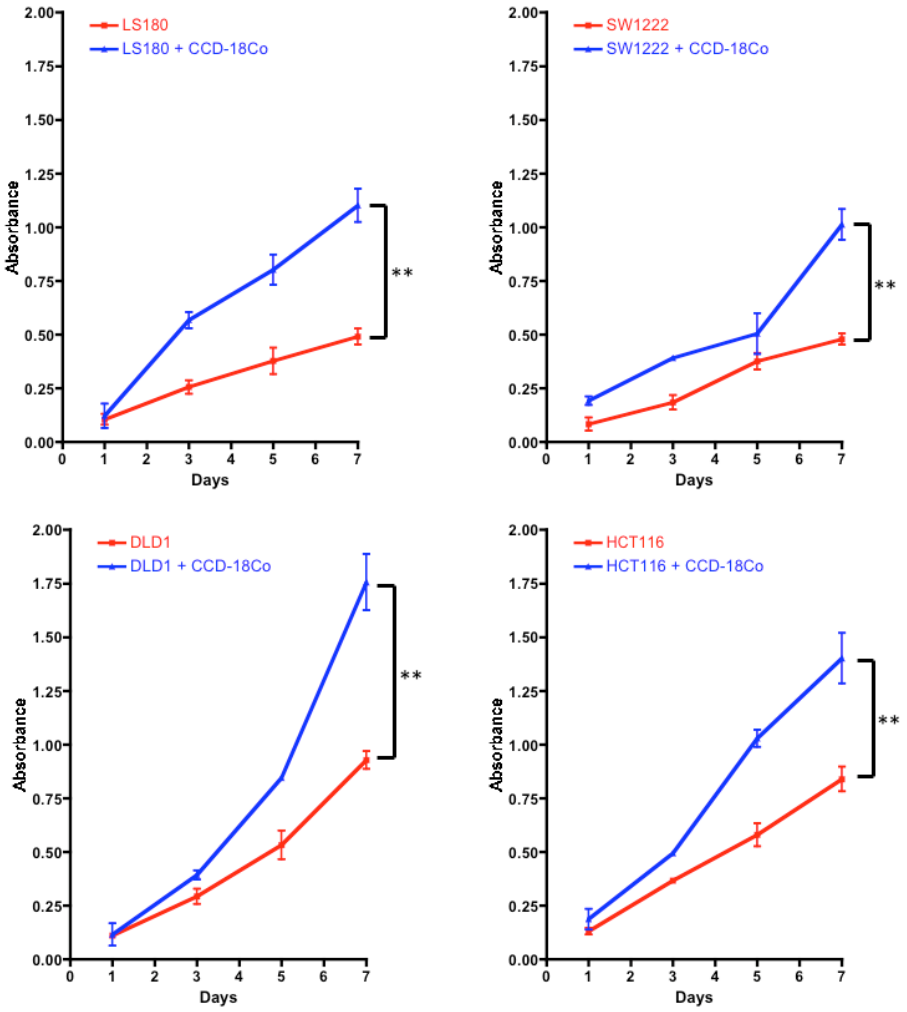


Figure 6.6 Growth curves corresponding to MTS-measured cell proliferation for co-cultured cells. Plotted points represent six data points (two experiments done in triplicate). For all four cell lines tested, the co-cultures were more proliferative and grew the most rapidly over a seven-day test period. The slopes of the sorted populations were compared to that of the unsorted, bulk, populations via Student's *t*-tests. * - 0.05 significance level, ** - 0.01 significance level, *** - 0.005 significance level.

Cell Line	Monoculture Slope (units/day)	Co-Culture Slope (units/day)
LS180	0.0640 ± 0.0030	0.1587 ± 0.0151
SW1222	0.0688 ± 0.0064	0.1288 ± 0.0298
DLD1	0.1344 ± 0.0173	0.2687 ± 0.0511
HCT116	0.1169 ± 0.0032	0.2090 ± 0.0146

Table 6.3 Summary of proliferation data from myofibroblast co-cultures transfections. Presented values represent six data points (two experiments done in triplicate). Among all cell lines, the co-cultures were the more proliferative and grew the most rapidly over a seven-day test period.

All four cell lines showed increases in proliferation upon co-culture with CCD-18Co myofibroblasts at a significance level of 0.01; indeed, the proliferation rates for all cell lines, as measured by the slopes of the growth curves, doubled in the presence of myofibroblasts (LS180 more than doubled its growth rate). Based upon these results, myofibroblast co-culture appears to promote proliferation of CSCs by virtue of general increases in cell proliferation.

6.2.4 Cell lines co-cultured with CCD-18Co cells exhibit differential CD55 expression profiles

The work of Huang and colleagues (2010) discovered that co-culturing cancer cells with fibroblasts can affect the CSC population by increasing its numbers—in this specific example, the population of $CD44^+CD24^-$ breast CSCs increased. To assess if CD55, the putative CSC marker previously identified, can be linked to the co-culture of myofibroblasts, epithelial cells were harvested, divided into 10^5-cell aliquots and combined with 0, 5 x 10^4, 10^5 or 2 x 10^5 myofibroblasts (producing myofibroblast : epithelial cell plating ratios of 0, 0.5, 1 and 2). Cell combinations were grown in tissue culture flasks for one week. For analysis, co-cultures were harvested, stained with anti-CD55 monoclonal antibody, counterstained with anti-EpCAM monoclonal antibody and analyzed via flow cytometry. To exclude the myofibroblasts from the CD55 expression studies (as they are $CD55^+$), EpCAM was used as a positive selector of the colonic epithelial cells; $EpCAM^-$ cells (the myofibroblasts) were gated out of the analysis (Figure 6.6). The resulting $EpCAM^+$ cells were assessed for CD55 expression (Figure 6.7), and the proportions of the culture that were positive for CD55 were quantified (Table 6.4).

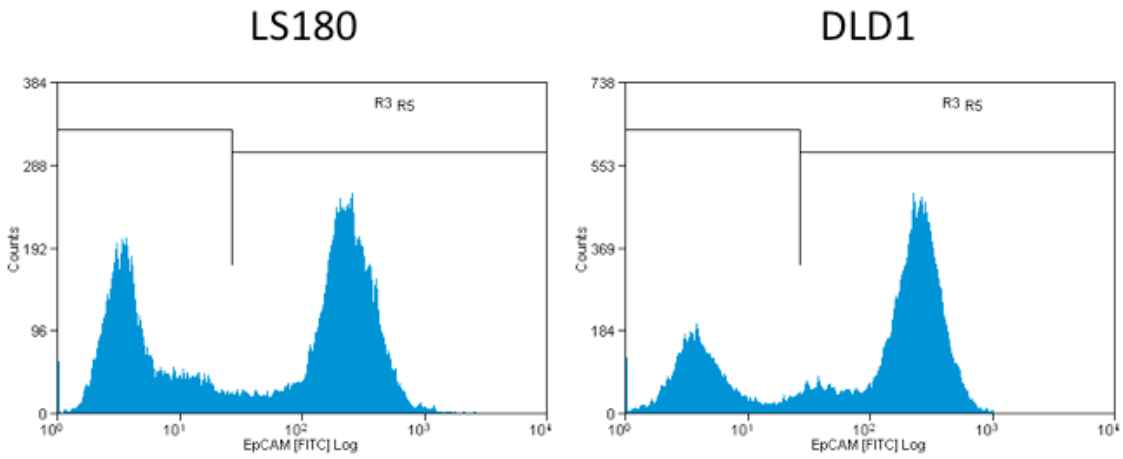


Figure 6.7 Example flow cytometry histograms of myofibroblast/epithelial cell co-cultures showing EpCAM expression in LS180 and DLD1. The myofibroblasts are EpCAM$^-$, while the epithelial cells are EpCAM$^+$. The positive population was included in the analysis of CD55 expression.

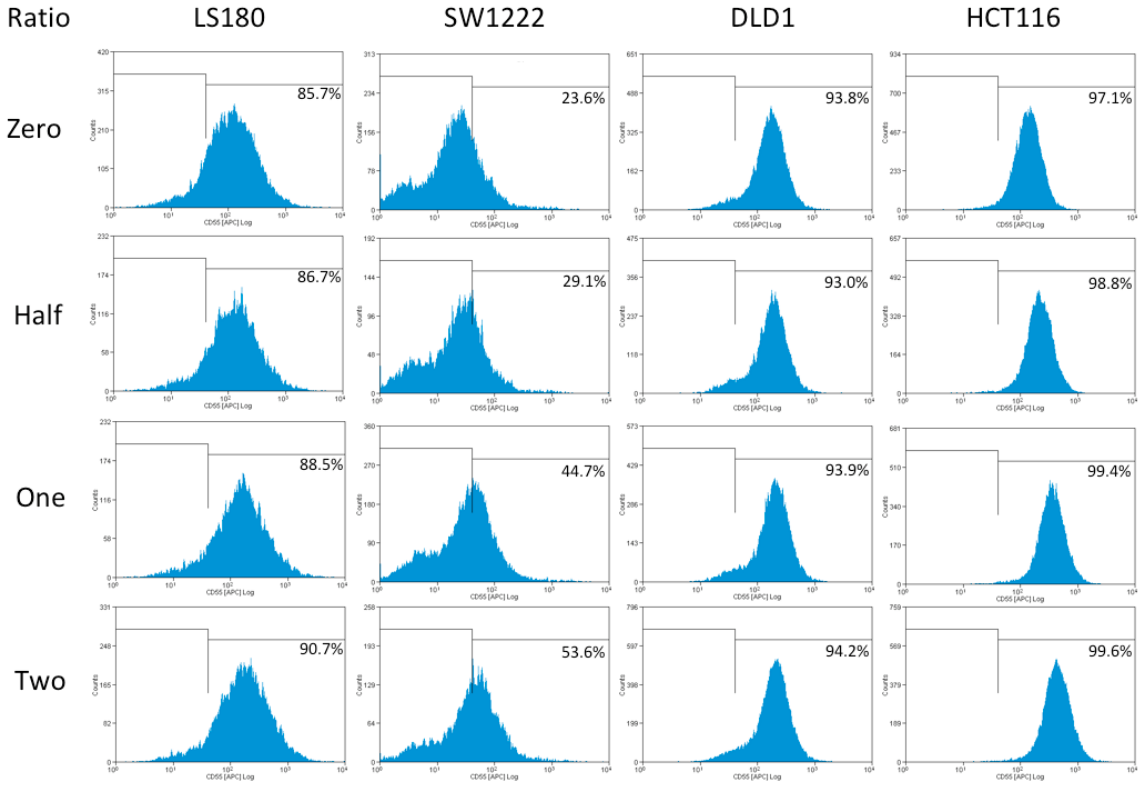


Figure 6.8 Flow cytometry histograms of CD55 expression for the four tested cell lines (LS180, SW1222, DLD1 and HCT116). The SW1222 cell line was sensitive to myofibroblast co-culture, and the proportion of cells expressing CD55 (as compared to the isotype control) more than doubled over the co-culture series. LS180 was relatively insensitive. Conversely, the two dense cell lines were not sensitive to co-culture; their populations of CD55$^+$ cells remained relatively constant, although the HCT116 cell line exhibited an overall increase in fluorescence level, observed by peak fluorescence, upon increase in myofibroblast number.

Cell Line	Myofibroblast : Epithelial Cell Ratio			
	0	0.5	1	2
	% Positive	% Positive	% Positive	% Positive
LS180	85.7	86.7	88.5	90.7
SW1222	23.6	29.1	44.7	53.6
DLD1	93.8	93.0	93.9	94.2
HCT116	97.1	98.8	99.4	99.6

Table 6.4 Summary of CD55 expression data from flow cytometry of myofibroblast co-cultures. Based upon the percentage of the distribution positive for CD55, SW1222 was sensitive to co-culture, demonstrating a doubling in the $CD55^+$ population as the number of co-cultured myofibroblasts increased. LS180 and the two dense lines were relatively insensitive to co-culture; the $CD55^+$ populations were essentially unchanged.

Co-culture of the CCD-18Co colonic myofibroblasts with SW1222 lumen cells led to an expansion of the $CD55^+$ CSC population in a dose-dependent manner—the proportion of cells expression CD55 rose dramatically from 23.6% to 53.6%, a doubling. The LS180 cell line was far less sensitive to myofibroblast-mediated induction of the $CD55^+$ population, while the two dense lines were similarly insensitive. This is likely a result of the high endogenous levels of CD55 expressed by LS180, DLD1 and HCT116, as measured by flow cytometry—nearly all cells from the three lines express the CD55 CSC marker, so it would have been unlikely to see further increases in an already ubiquitous cell population. However, slight increases in the peak fluorescence level of the CD55 distributions within LS180, DLD1 and HCT116 were observed—HCT116, in particular, saw an increase from approximately 150 to 500 fluorescence units, indicating that some induction of CD55 expression, albeit from positive levels to *very* positive levels, could have occurred. Nevertheless, for SW1222, CCD-18Co co-culture did induce proliferation of the $CD55^+$ CSC population.

6.3 Discussion

Cancer studies attempting to understand the environment in which tumor cells reside have often used co-culture experiments of the cancerous epithelial cells with associated fibroblasts or myofibroblasts to understand the interplay between the two cell types, among the many cell types within the tumor niche. However, the roles of local non-epithelial cells in regulating CSCs have not been studied extensively. Of course, mouse xenograft experiments have been conducted for many years, and the sites in which tumor stem cells are injected could be thought of as nascent tumor niches in which other, non-cancerous tissues could exert effects upon the CSCs. However, xenografts are generally not used within this context—their primary purpose is to serve as an *in vivo*-like "incubator" of cancer stem cells, and the specific cells located near the tumor cell injection site are rarely elucidated.

Thus, to understand the effects of these niche-associated cells on the cancer stem cell compartment in a panel of colorectal cancer cell lines, four cell lines (two lumen, differentiated lines and two dense, undifferentiated lines) were selected and co-cultured under a variety of conditions with myofibroblasts. In three-dimensional growth matrix conditions in the presence of calcium-fortified growth media, the addition of myofibroblasts triggered increased clonogenicity and, in the case of the lumen cell lines, increased lumen formation. In two-dimensional conditions, the addition of myofibroblasts appeared to have triggered increased proliferation and an expansion of the $CD55^+$ cancer stem cell population in the SW1222 cell line, with the other three cell lines exhibiting marginal responses to the myofibroblast additions to the culture.

The clonogenicity and proliferation results presented showed that myofibroblast co-culture promotes the expansion of the CSC population by virtue of higher clone-

formation and growth rates. Surprisingly, the clonogenicity regression results presented showed that the undifferentiated dense lines were more sensitive to clone-formation induction than the differentiated lumen lines. This could be a result of a higher sensitivity to the pro-stemness signals emitted by the myofibroblasts—the differentiation-sensitive lumen lines may not be able to respond as strongly to such signals from the myofibroblasts. Interestingly, the proportions of lumens increased as well between the two lumen lines. This is likely linked to the increases in CSC population—the dense lines, by virtue of their inability to give rise to lumen-containing colonies, are unable to respond to such signals, and thus do not generate any lumen colonies upon myofibroblast co-culture. Lumen lines, conversely, can respond to such environment signals (such as those found in the serum of the culture medium used)—an expansion of the CSC population in lumen lines would yield more cells that are capable of forming lumens. A potential future experiment to test this hypothesis would be to co-culture epithelial cells and myofibroblasts in serum-free, stemness-promoting medium—if the increases in lumen-forming colonies are attributable to an expansion of CSCs that differentiate in serum media, such differentiation should disappear in the serum-free media, leaving just increases in total clonogenicity.

In the end though, the results presented here support the notion that myofibroblasts drive pro-clonogenic, pro-proliferative and pro-CSC pathways within colorectal cancer cell lines. The CCD-18Co cell line possesses an ability to promote the CSC population.

CHAPTER 7

GENERAL DISCUSSION AND FUTURE WORK

7.1 Colony morphology in three-dimensional culture is linked to differential gene expression

Although cancer cell lines are generally grown *in vitro* as adherent cultures in plastic flasks, such culture conditions are quite foreign to the *in vivo* conditions experienced by tumor cells. Cells are in contact with each other in three dimensions and are surrounded by an extracellular matrix (ECM) or other protein scaffold. Growth factors diffuse into the matrix to reach cell surface receptors, and cells organize into larger structures.

Experiments presented here show that cell lines grown in an ECM-like substrate (a Matrigel™ ECM formulation) possess the capacity to give rise to three-dimensional colonies with distinctive morphologies. Although each cell line formed *in vitro* colonies differently, and the resulting colonies are organized differently (likely a result of the unique properties of each primary tumor from which the cell lines were derived), striking similarities in the cell lines' colony morphologies and in the relatively homogeneous nature of morphology within individual lines allowed for the classification of colonies, and cell lines, into four groupings, viz. network, lumen, dense and intermediate. Each classification was characterized by the organization and polarity of cells within a colony.

The most interesting grouping was the lumen group, in that the cell lines visually appeared to possess an ability to give rise to colonies that spontaneously differentiate into the columnar (which stain positive for CDX1), enteroendocrine (which stain positive for chromogranin A) and goblet (which stain positive for mucin) cell lineages, in varying proportions as found in the human gut (Yeung, et al., 2010). Confirmatory immunohistochemistry experiments on the lumen-containing colonies to determine the presence and relative proportions of the three cell lineages would have been preferable in order to corroborate the findings of Yeung and colleagues. Furthermore, the lumen

colonies possessed a central open space around which cells visually appeared to be organized in a manner very reminiscent of the mature colonic epithelium. Further work, such as immunohistochemistry of colony sections and electron microscopy of individual cells at the apical and basolateral surfaces of the colonies, should be performed to characterize the presence and nature of such polarized cells. Staining for the presence of brush borders on the apical surface and for the presence of E-cadherin junctions and β-catenin near the basolateral surface would be useful in confirming the polarized nature of the lumen colonies.

The similarities among colony morphologies within each classification implied that there were likely to be common gene regulatory patterns among cell lines of a particular organization. Consequently, previously collected gene expression data for the nine lumen lines and the seven dense lines (which appeared undifferentiated) were compared to each other to identify the transcriptional sources of this difference in colony morphology. Of the 428 probesets found to be significantly upregulated in the lumen lines, a preponderance of them have been previously associated with differentiation or the dampening of colorectal carcinogenesis. These include *FABP1* (Lawrie, et al., 2004), *MUC3A* (Ho, et al., 1993), *TFF3* (Babyatsky, et al., 2009), *VIL1* (Yamamichi, et al., 2009) and *CDX1* (Chan, et al., 2009). Of the 394 probesets found to be upregulated in the dense lines, many have been found to have stemness- and tumor growth-promoting properties, including *RUNX2* (Wai, et al., 2006), *FGF2* (Navarro-Alvarez, et al., 2008), *BMI1* (Molofsky, et al., 2003) and *CD55* (Xu, et al., 2007).

Therefore, the dense cell lines appear to be made up of cells that primarily behave and express genes consistent with stem cell populations that do not differentiate, while the lumen cell lines appear to be made up of cells that can differentiate. To understand the

control of colorectal cancer stemness and differentiation, the results of the cell line panel indicate that the characteristics of dense lines can be studied as surrogates for characteristics of cancer stem cells (CSCs), while those of lumen lines can be studied as models of cancer differentiated cells.

The idea that cells grown in long-term *in vitro* culture still possess the ability to form differential colony morphologies in three-dimensions is not novel. Lang and colleagues (2001) grew the normal PNT2-C2 and cancerous PC-3 prostate cell lines in Matrigel™ and found that both were able to give rise to colonies, with the PNT2-C2 cells able to differentiate *in vitro*. Richman and Bodmer (1988) found that the SW1222 colon cancer cell line could similarly differentiate in collagen I. Kenny and colleagues (2007) took this approach one step further and assayed a 24-cell line breast cancer panel to identify genes whose expression could be associated with each of four colony morphologies listed; computational methodology was devised to automate and the analysis of breast cancer colony morphology and gene expression (Han, et al., 2010) to highlight individual genes associated with individual colony morphologies, such as PPARγ and stellate cells. Within prostate cancer, Härma and colleagues (2010) correlated colony morphology to drug responses, invasiveness and gene expression in a 29-cell line prostate cancer panel. However, few have attempted to link colony morphology to an exogenous characteristic, such as the presence or absence of stem or differentiated cells within the cell line culture, or have attempted to link the gene expression profiles associated with colony morphology to such a characteristic. The results presented here, thus, are unique for colorectal cancer. To identify novel markers of cancer stemness or to understand the genes involved in triggering cancer cell differentiation, the genes upregulated in the respective cell line groupings can serve as useful starting points. As such, the remainder of the experiments

presented here focused on three highlighted genes from the comparisons: *BMI1*, *CD55* and *CDX1*.

The results of microarray analyses are often used as training sets in order to make predictions of biological or clinical significance based upon sample gene expression data. To date, the colony morphology studies done on cancer cell lines (Kenny, et al., 2007; Härma, et al., 2010) have restricted the associations between gene expression and morphology to a training set. To test the true power of the gene associations uncovered by the experiments presented here, a second cell line panel (a prediction set) can be assembled and tested for colony morphology. If the gene associations identified in the training set, such as CDX1 with lumen lines and CD55 with dense lines, are rigorous, their expression patterns within the prediction set should be predictive of colony morphology. The recent completion of additional human genome microarrays on an additional fifty cell lines within the CIL will allow such experiments to take place. Alternatively, the entire panel of eighty cell lines can be combined to offer increased statistical significance of differentially-expressed genes and to allow for greater exclusion of spurious probesets.

An additional avenue of further experimental could focus on the first colony morphology identified, the network class. Compared to the other three classes, the network classification was the most radically different—the other three morphologies formed semi-spherical colonies, while the network morphology did not. The fibrous, spindle-shaped cells that grew out of the Colo741, LS174T and RKO cell lines are reminiscent of the morphological appearance of epithelial cells undergoing an epithelial-mesenchymal transition (EMT), in which epithelial cells undergo a transformative process that liberates them from well-organized epithelial layers (Agiostratidou, et al., 2007) to form invasive,

migratory cells similar in behavior to mesenchymal cells. Indeed, one of the hallmarks of EMT, a switch in expression of the cell-to-cell cadherin adhesion proteins from E-cadherin to N-, P- or R-cadherin (Makrillia, et al., 2009; Maeda, et al., 2005), was identified in the microarray analysis—all three network cell lines are negative for E-cadherin and possess methylated *CDH1* promoter regions (Dikomitou, 2009) that silence *CDH1* transcription. Such a downregulation of E-cadherin has been associated with increased invasiveness in colorectal cancer cells (Kinsella, et al., 1994). However, other cell lines, such as NCIH716, are also negative for E-cadherin (Bracht, et al., 2010) or possess methylated *CDH1* promoter regions (Dikomitou, 2009). These lines form colonies that exhibit at least some cell-to-cell adhesion. Therefore, the existence of the network colony morphology cannot easily be explained by the dysregulation of a specific protein. Nevertheless, the link to E-cadherin becomes important, particularly when examining the morphological differences between the intermediate LS174T line and the lumen LS180 line; both cell lines are derived from the same primary tumor (Rutzky, et al., 1980) but differ only on E-cadherin expression. LS174T is negative, while LS180 is positive (Kinsella, et al., 1994). Identification of other factors influencing the formation of the network morphology, such as presence or lack of other cadherin proteins, and the extent to which these factors are or are not related to EMT would prove interesting.

7.2 CD55 is a marker of cancer stemness

The morphology microarray expression analysis identified CD55 as a potential marker of cancer stemness by virtue of its association with the undifferentiated dense cell line morphology. Since CSCs are characterized by their ability to self-renew, differentiate and

initiate tumors (Clarke, et al., 2006), in vitro assays measuring these characteristics would be necessary to assess the usefulness of CD55.

Flow cytometry analysis of CD55 expression in a panel of six cell lines found variable levels of expression, with four (DLD1, HCA7, HCT116 and LS180) expressing high levels of the protein. Although it is thought that CSCs must be rare (likely a consequence of the hematological system in which much of the early work in characterizing CSCs was conducted), recent evidence (Quintana, et al., 2009; Kelly, et al., 2007) has challenged this dogma. Indeed, mathematical modeling of the colonic crypt by Johnston and colleagues (2010) has shown that the stem cell compartment need not be rare. Although Kai and colleagues (2009) posited that a large population of tumor-initiating cells in the HCT116 colorectal cancer cell line implies that the stochastic model of carcinogenesis is more correct, evidence presented here and elsewhere (Yeung, et al., 2010) suggests that differences in CSC capabilities can be identified within cell populations that most probably to possess a large proportion of CSCs. Thus, the result that many cell lines express CD55 at near-ubiquitous levels should not discount the notion of CD55 as a marker of cancer stemness—as the results showed, distinct subpopulations within the $CD55^{+}$ population possessed different behavioral and biological characteristics.

Culturing $CD55^{high}$ cells in serum-free medium, known to promote the retention of the CSC phenotype (Ricci-Vitiani, et al., 2007), and in serum medium, known to promote cell differentiation (Vermeulen, et al., 2010), revealed the population to possess self-renewal and differentiation capabilities. *In vitro* clonogenic assays in the ECM-like substrate Matrigel™ showed $CD55^{high}$ cells to initiate more tumor colonies, while MTS proliferative assays showed them to possess a higher growth potential. Interestingly, similar to the results of Harper and colleagues (2010), the $CD55^{high}$ putative CSCs were

preferentially in the G_2/M cell cycle phase, implying a greater resistance to apoptosis. Finally, $CD55^{high}$ cells expressed higher levels of BMI1 and lower levels of CDX1 as compared to the bulk culture, again characterizing the $CD55^{high}$ population as stem-like (Sangiorgi and Capecchi, 2008) and undifferentiated (Chan, et al., 2009).

However, CD55 is but one marker of cancer stemness—it is unlikely that the sole mediator of the CSC phenotype is a single protein. The clonogenicity and proliferation of $CD55^{high}$ cells was increased by no more than two-fold compared to the bulk culture, and $CD55^{low}$ cells still possessed some clonogenic and proliferative activity. Indeed, Stuelten and colleagues (2010) assessed the ubiquity of eight different putative CSC markers (ALDH1, CD14, CD24, CD44, CD133, CD166 and PgP) in the NCI-60 panel of cancer cell lines, only to find extreme heterogeneity among the cells vis-à-vis marker expression. Thus, it is far more likely that a multiple selection, in which two or more markers are used to purify the CSC population, will yield a culture of cells far more enriched with cancer stem cells. Based upon previous work in the CIL linking the $CD24^{high}CD44^{high}$ population to cancer stemness (Yeung, et al., 2010), CD55 was added to the extant dual-sort to create a simultaneous three-color fluorescence-activated cell sort. The addition of CD55 further enhanced clonogenicity and proliferation—further evidence linking CD55 to cancer stemness. In addition, the enriched population possessed a higher propensity to differentiate into lumens, indicating that a larger proportion of the population did in fact possess the ability to differentiate, a hallmark of cancer stemness. The methods outlined here were ideal in identifying CD55 as a potent marker of cancer stemness—cell lines offer large amounts of homogeneous, well-characterized (through mutation and expression studies) material for multiple experiments, while *in vitro* growth studies allow for the study of putative CSCs in a well-controlled environment. Of course, many studies of CSCs utilize cells from primary tumor biopsies, test cell tumorigenicity in mouse

xenograft models and assess the chemoresistance of isolated cells—these experiments should be performed to validate CD55's role as a CSC marker.

Preliminary mouse subcutaneous xenograft experiments comparing the $CD24^{high}CD44^{high}CD55^{high}$ subpopulation with the $CD24^{low}CD44^{low}CD55^{low}$ within the SW1222 cell line have shown the high population to generate far larger tumors (n = 4, mean of 437 mm^2) than the low population (n = 4, mean of 50 mm^2) at a significance level of less than 0.001 (Trevor Yeung, unpublished results). While the xenograft was heterotopic for ease of transplantation and tumor monitoring, the cellular environment in which the xenograft was placed was that of the hypodermis, not of the colon. The growth factor signals, stromal cells and extracellular matrix composition of the hypodermis would all differ from those of the colon and would not reflect the true milieu of colorectal tumors. Renal capsule xenografts, a less heterotopic alternative used by O'Brien and colleagues (2007) to characterize $CD133^+$ colon CSCs, resemble the colon more closely and, thus, would likely generate results more closely mirroring the true *in vivo* behavior of colon CSCs. However, a better orthotopic alternative for xenograft experiments would be for CSCs to be injected into a site even more similar to the colon, such as the wall of the cecum. As demonstrated by Céspedes and colleagues (2007) with the colon cancer cell lines DLD1, HCT116 and SW620, cecal xenografts recapitulate the environment of the colon very closely and allow for the modeling of downstream tumor behaviors, such as organ wall invasion and metastasis.

The clone-formation rates achieved by the $CD24^{high}CD44^{high}CD55^{high}$ as a proportion of the total cells plated was nowhere near unity—HCT116 $CD24^{high}CD44^{high}CD55^{high}$ cells were only able to generate clones 21.9% of the time. To isolate a cell population that could generate clones 100% of the time, a pure CSC population, the three-color cell sort

could be extended to encompass more markers of cancer stemness—indeed, eight-color flow cytometry experiments are routinely done in the immunology and hematology fields. Although CD133, one of the first cell surface markers to be linked to stem cells in bowel cancer (Ricci-Vitiani, et al., 2007; O'Brien, et al., 2007), could serve as an additional marker to the extant CD24/CD44/CD55 methodology, recent evidence (Navarro-Alvarez, et al., 2010; Kemper, et al., 2010; Shmelkov, et al., 2008) has cast in doubt the usefulness of CD133. Alternatively, a more interesting marker is aldehyde dehydrogenase-1 (ALDH1). Characterized as a marker of cancer stemness in breast (Ginestier, et al., 2007), lung (Jiang, et al., 2009), head and neck (Chen, et al., 2009) and colorectal (Huang, et al., 2009), ALDH1 has been associated with higher tumor grades in colorectal cancer (Lugli, et al., 2010) and is thought to be a detoxifying enzyme that inactivates chemical insults, such as the chemotherapy drug cyclophosphamide (Dylla, et al., 2008). Furthermore, the colony morphology microarray analysis identified an ALDH1 family gene, *ALDH1A3*, which was significantly upregulated in the undifferentiated dense colony cell lines. With the development of the ALDEFLUOR® enzyme assay, the expression of the intracellular aldehyde dehydrogenase can be probed externally via flow cytometry by tracking the metabolism of a fluorescent substrate, negating the need for cellular permeabilization (Tan and Lee, 2009). Sorting cells on the basis of CD24, CD44, CD55 and ALDH1 might not only increase clonogenicity even more, but also further enrich the isolated population for cells that are capable of differentiating into lumen-containing colonies.

Another approach in extending the CD55 findings would be to ascertain the functional relevance of CD55 expression in colorectal cancer cells. The CSC field is littered with potential markers for which a correlative link, but not a causal link, has been found to cancer stemness. For example, extant evidence linking CD133 to the CSC phenotype has

been primarily correlations. The biological significance of CD133, such as a ligand or downstream signaling interaction, has not been identified (Horst, et al., 2009). Reduced CD133 expression has been associated with increased colon cancer cell differentiation (Feng, et al., 2010), but siRNA knockdown of CD133 did not induce such differentiation. RNA interference of CD133 also does not lead to a reduction in cell proliferation or increased apoptosis (Elsaba, et al., 2010). Conversely, the CSC marker CD44 has been shown to possess functional relevance in colorectal carcinoma, in addition to its relevance as a CSC marker—it acts as an adhesion molecule that binds to extracellular components of other cells and of the ECM and thus, mediates cell motility, among other characteristics (Ponta, et al., 2003). Within colorectal cancer, siRNA knockdown of CD44 expression severely dampened clonogenicity and tumor initiation in a mouse xenograft model (Du, et al., 2008), while construction of a stable CD44 siRNA-transfected HT29 cell line found reduced ECM adhesion, increased sensitivity to induced apoptosis, dampened clonogenicity and suppressed tumor formation in mice (Subramaniam, et al., 2007) among such cells.

Induction of *CD55* expression through cDNA transfection or ablation of *CD55* expression through siRNA knockdown would allow for an understanding of the role, if any, of CD55 in promoting clonogenicity, tumorigenicity and chemoresistance—if CD55 were of functional relevance to the CSC phenotype and is necessary for the CSC compartment to survive in colon cancer, then cDNA induction experiments should show an enhancement of clonogenicity, tumorigenicity and chemoresistance (and vice versa for siRNA knockdowns). Furthermore, assessing the chemoresistance of $CD55^{high}$ cells as compared to the bulk population through drug screens would allow for a more clinical understanding of CD55's role in cancer stemness. The long-term goal of CSC research is to develop new cancer therapies that are more targeted and effective than current

treatments—any functional experiments focusing on CD55 should include an assessment of the relative resistance or sensitivity of $CD55^{high}$ CSCs to different chemotherapeutic (and radiotherapeutic) modalities as compared to bulk cancer cells.

Further studies should also focus on the role of CD55 in patient tumors. Durrant and colleagues (2003) found an association between CD55 immunohistochemical staining intensity in patient tumor biopsies and poor patient prognosis. Within the context of CD55 as a marker of cancer stemness, the findings of Durrant should be extended to assess any associations between CD55 biochemical findings in patients, such as tumor staining patterns, specific tumor mRNA and protein expression levels and presence of CD55 on any circulating tumor cells, and clinical findings, such as tumor stage, patient prognosis, overall survival and response to treatment.

7.3 BMI1 promotes cancer stemness, while CDX1 promotes sensitivity to cancer differentiation

The morphology microarray expression analysis also identified BMI1 as a potential marker of cancer stemness and CDX1 as a potential marker of cancer differentiation. While the cancer stem cell role of CD55 had not been previously elucidated, BMI1 and CDX1 have been characterized within the contexts of stem cell biology. The mRNA expression of *BMI1* is heightened in tumorigenic breast cancer spheroids—induction of *BMI1* expression leads to increased spheroid generation, while expression knockdown leads to decreased spheroid generation (Liu, et al., 2006). CDX1 directly regulates *KRT20* expression in the colon and is expressed ubiquitously in the differentiated luminal region of the colonic crypt (Chan, et al., 2009). However, their roles in mediating CSCs in colorectal cancer have not been studied.

To do so, two cDNA plasmid constructs that constitutively drive expression of *BMI1* or *CDX1* as well as a reporter green fluorescent protein were synthesized and transiently transfected in a panel of four cell lines, two lumen and two dense—Western blotting and fluorescence imaging confirmed that the two cDNAs were expressed highly and that such expression could persist for up to seven days. Although the cells were cultured for more than seven day (two weeks), the initiation and initial differentiation (if any) of a colony likely occur only hours or days after the start of the culture—even if gene expression dropped off in the ensuing two weeks, the initial steps toward the generation of a clone were already in motion. *BMI1*-transfected cells grown in Matrigel™ showed increases of 30-50% in clonogenicity, both in lumen and dense lines. However, a more interesting result was that the proportions of colonies possessing lumens decreased upon transfection—BMI1 appeared to have dampened differentiation within the lumen lines and promoted the maintenance of a stem-like phenotype, a causal link between BMI1 and stemness, not only a correlation, as has been found by Kirkland (2009). *CDX1*-transfected cells grown in Matrigel™ showed opposite effects; clonogenicity was reduced by 50-60% in all four cell lines. In addition, *CDX1* induction promoted the expansion of colonies containing lumens, both in lumen lines but also in the undifferentiated, non-lumen-forming dense lines. CDX1 appeared to have sensitized cells to environment differentiation cues, implying a causal link between CDX1 and differentiation. *BMI1*-transfection promoted cell proliferation, while *CDX1*-transfection dampened cell proliferation.

Most cancer stem cell markers have been identified by their abilities to purify tumorigenic, self-renewing subpopulations from a cell culture—their functional roles are later characterized. To this end, a reporter plasmid construct was synthesized incorporating a green fluorescent protein reporter under transcriptional control of an

1154-base upstream promoter region of CDX1. Transfection of the plasmid labeled $CDX1^+$ cells by fluorescently-marking cells possessing active CDX1 promoter regions and, thus, active CDX1. Flow cytometric comparison of $CDX1^+$ cells to the bulk culture identified a correlation between high CDX1 expression and low CD55 expression, establishing for the first time the role of CDX1 as a negative regulator of CSCs. The results of these experiments, therefore, confirm the critical roles of BMI1 and CDX1 in the regulation of colorectal cancer stem and differentiated cells.

The success of the CDX1 promoter construct in identifying a $CD55^{low}$ population easily lends itself to potential fluorescence-activated cell sorts to select for expression, or lack of expression, of intracellular CSC markers. The findings with the construct can be extended to the purification of $CDX1^+$ cells for clonogenicity, proliferation, xenograft tumorigenicity, drug response and many other experiments to ascertain the usefulness of CDX1 as a negative marker of cancer stemness. Furthermore, transfection of the construct can be combined with flow cytometric staining for other markers of cancer stemness, including CD24, CD44 and CD55, in order to purify a population that is positively-selected for stemness and negatively-selected for differentiation. While novel, selecting for cancer stem cells through promoter-reporter assays is not without precedent. Levings and colleagues (2009) constructed an Oct-4 promoter vector to identify an $Oct\text{-}4^+$ population in the OS521 osteosarcoma cell line; the fluorescent population was positive for CD105 and ICAM-1, while the bulk population was heterogeneous in expression. Furthermore, $Oct\text{-}4^+$ cells were able to generate more tumors in a mouse xenograft model than $Oct\text{-}4^-$ or bulk cells. Tamase and colleagues (2009) performed similar experiments in brain tumor cells using a nucleostemin promoter-reporter assay.

Stable transfection of the CDX1 promoter construct and a (potential) BMI1 promoter construct into cell lines can also allow for a detailed and time-sensitive assessment of activation of the two genes as cells divide and grow. Yeung and colleagues (2010) found in lumen-containing colonies that CDX1 expression was concentrated near the periphery of colonies, akin to the luminal surface expression that Chan and colleagues (2009) found in the colonic epithelial tissue. The staining pattern of BMI1 is vaguer—while it is not appreciably expressed in normal mucosa, it is expressed widely in colorectal tumors (Kim, et al., 2004). However, the mechanism by which these expression patterns emerge is largely unknown. Cells expressing stably-transfected CDX1 and BMI1 promoter vectors can be grown and photographed over time to assess changes in CDX1 and BMI1 expression as colonies, or tumors (in mice) grow. Cells treated with exogenous factors, such as growth factors or cytotoxic drugs, can be observed in real-time to understand temporal changes in CDX1-mediated differentiation and/or BMI1-mediated stemness.

7.4 Colonic myofibroblast co-culture promotes cancer stemness

The colony morphology experiments were initially done in order to understand the growth and organizational behavior of cells when grown in conditions that resemble the *in vivo* environment. However, the *in vivo* environment not only contains three-dimensional ECM scaffolds but also non-epithelial cells, such as intestinal myofibroblasts, that secrete growth factors and cytokines (Humphries and Wright, 2008) that are likely necessary to maintain the stem cell population within its niche.

To assess the impacts of myofibroblasts on the maintenance of the stem cell phenotype, a panel of four cell lines, two lumen and two dense, were grown in Matrigel™ in direct

contact with varying ratios of CCD-18Co myofibroblasts, from none to double the number of epithelial cells. Co-cultured cells showed increases of up to 50% in clonogenicity, both in lumen and dense lines—the lumen lines showed progressive increases in clone formation as the number of myofibroblasts increased. Interestingly, the proportions of colonies possessing lumens increased upon co-culture, implying a causal link between the addition of stem cell niche cells and the survival and expansion of the CSC population. Cancer cell proliferation additionally increased upon addition of myofibroblasts. Flow cytometric analysis of cancer cells cultured alongside myofibroblasts revealed a positive correlation between the number of myofibroblasts and proportion of epithelial cells that were positive for the cancer stem cell marker CD55. However, this effect was only pronounced in the SW1222 cell line; the dense lines were less sensitive to the CD55-promoting effects of the CCD-18Co cells, likely a result of already-undifferentiated nature of the dense cell lines. In total, the addition of the myofibroblasts promoted the expansion of CSCs, as was posited by recent work by Vermeulen and colleagues (2010), a conclusion that can easily be validated with the co-culture of primary colonic myofibroblasts with cancer cells to assess *in vitro* clonogenicity and the co-injection of myofibroblasts with cancers into immunocompromised mice to assess *in vivo* tumorigenicity.

The histological architecture of the colonic crypt is such that the myofibroblasts reside inferior to the epithelial cells (Humphries and Wright, 2008); the epithelial tissue layer is polarized in this aspect. *In vitro* models, such as those of Ootani and colleagues (2009), have succeeded in culturing large intestine cells by growing epithelial cells superior to a myofibroblast feeder layer within collagen I scaffolds, separated by a permeable membrane. The success of the model largely depended on the inclusion of a secretory

myofibroblast layer (Ootani, et al., 2009) that could maintain the stem cell niche within the superior epithelial cells, allowing for sustained growth of the culture over many months. This could also be the case for colorectal cancer cell lines. To complement the experiments reported in this thesis involving the direct *in situ* co-culture of epithelial cells and myofibroblasts, cancer cells should be cultured within thin layers of Matrigel™ overlaid on a myofibroblast feeder later or cultured on the surface of such thin layers. Experiments involving the separation of the myofibroblasts from the epithelial cells via an air-liquid interface and permeable membrane, similar to that reported by Ootani and colleagues, or via a Transwell™ insert, would also be of interest. Such separation studies would allow for a greater understanding of the role of direct epithelial cell-myofibroblast contact versus the role of myofibroblast soluble factor secretion on the proliferation of colorectal cancer cells.

It is also clear that the secretion of growth factors mediates the effects of myofibroblasts in maintaining the stem cell niche and cytokines that influence epithelial cell function (Yen and Wright, 2006). Work by Vermeulen and colleagues (2010) attempted to identify the secreted factors that may be responsible for promoting the stem cell population within colorectal cancer. A cytokine assay revealed hepatocyte growth factor (HGF) to be an abundant secreted factor in myofibroblast-conditioned media, and that secretion of HGF may be associated with activation of the Wnt pathway in colorectal CSCs. However, the conditioned media were derived from serum-free stem cell media and lacked any exogenous growth factors that allow cultured cells to differentiate. Thus, while HGF was secreted, it may not play a major role in promoting the stem cell niche and/or regulating the orderly differentiation of CSCs into differentiated cells, as was observed in the myofibroblast co-culture experiments described here. Microarray analyses of

myofibroblast cultures, including CCD-18Co, may prove useful in identifying specific secreted proteins that are highly-expressed in such cells, while antibody arrays of myofibroblast-conditioned media and ECM (as growth factors often bind to the ECM) containing some serum could prove more useful in identifying factors that regulate the stem cell population. Colony experiments involving varying concentrations of the identified growth factors, in varying combinations, may prove fruitful in ascertaining the specific cocktail of myofibroblast-secreted proteins and lipids that allows for the proper control of cancer stem cells in *in vivo*-like conditions.

7.5 Conclusion

To summarize, the experiments reported herein have identified a link between cell line colony morphology and differentiation status, and that such status can be utilized to identify novel markers of cancer stemness and controllers of cancer differentiation. One of these identified markers, CD55, preferentially purifies a subpopulation of colorectal cancer cells that can not only self-renew and differentiate, but also initiate colonies and proliferate at rates far higher than the bulk population. $CD55^{high}$ cells also express high levels of the stemness-associated marker BMI1 and low levels of the differentiation-associated marker CDX1, further confirming the role of CD55 in marking CSCs. Inducing BMI1 and CDX1 expression increases and decreases the CSC population, respectively, while labeling $CDX1^{+}$ cells in situ via a *CDX1* promoter fluorescent promoter vector can identify a CSC-poor subpopulation. Finally, recapitulating the *in vivo* stem cell niche through epithelial cell-myofibroblast co-culture promote the CSC population, as measured by CD55 expression, and the CSC phenotype, as measured by clonogenicity and proliferation. Clearly, studying the regulation of BMI1, CD55 and

CDX1 within the context of cancer stemness and differentiation proved and should continue to prove useful in understanding the mechanisms of colorectal carcinogenesis and tumor proliferation.

CHAPTER 8

REFERENCES

1. Abdouh M, Facchino S, Chatoo W, Balasingam V, Ferreira J, Bernier G. BMI1 sustains human glioblastoma multiforme stem cell renewal. J Neurosci. 2009; 29 (28): 8884-96.

2. Agiostratidou G, Hulit J, Phillips GR, Hazan RB. Differential cadherin expression: potential markers for epithelial to mesenchymal transformation during tumor progression. J Mammary Gland Biol Neoplasia. 2007; 12 (2-3): 127-33.

3. Aigner F, Fritsch H. Anorectal and colonic anatomy. In: Givel J-C, Mortensen NJ, Roche B, editors. Anorectal and Colonic Diseases: A Practical Guide to their Management. 3rd ed. Heidelberg: Springer-Verlag; 2010. p. 3-19.

4. Al-Assar O, Muschel RJ, Mantoni TS, McKenna WG, Brunner TB. Radiation response of cancer stem-like cells from established human cell lines after sorting for surface markers. Int J Radiat Oncol Biol Phys. 2009; 75 (4): 1216-25.

5. 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.

6. Alkema MJ, Wiegant J, Raap AK, Berns A, van Lohuizen M. Characterization and chromosomal localization of the human proto-oncogene BMI-1. Hum Mol Genet. 1993; 2 (10): 1597-603.

7. Andoh A, Bamba S, Fujiyama Y, Brittan M, Wright NA. Colonic subepithelial myofibroblasts in mucosal inflammation and repair: contribution of bone marrow-derived stem cells to the gut regenerative response. J Gastroenterol. 2005; 40 (12): 1089-99.

8. Andoh A, Shimada M, Araki Y, Fujiyama Y, Bamba T. Sodium butyrate enhances complement-mediated cell injury via down-regulation of decay-accelerating factor expression in colonic cancer cells. Cancer Immunol Immunother. 2002; 50 (12): 663-72.

9. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer. 1954; 8 (1): 1-12.

10. Aulmann S, Waldburger N, Penzel R, Andrulis M, Schirmacher P, Sinn HP. Reduction of CD44(+)/CD24(-) breast cancer cells by conventional cytotoxic chemotherapy. Hum Pathol. 2010; 41 (4): 574-81.

11. Babyatsky M, Lin J, Yio X, Chen A, Zhang JY, Zheng Y, et al. Trefoil factor-3 expression in human colon cancer liver metastasis. Clin Exp Metastasis. 2009; 26 (2): 143-51.

12. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature. 2009; 457 (7229): 608-11.

13. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007; 449 (7165): 1003-7.

14. Bartkowiak D, Stempfhuber M, Wiegel T, Bottke D. Radiation- and chemoinduced multidrug resistance in colon carcinoma cells. Strahlenther Onkol. 2009; 185 (12): 815-20.

15. Besnard P, Niot I, Poirier H, Clement L, Bernard A. New insights into the fatty acid-binding protein (FABP) family in the small intestine. Mol Cell Biochem. 2002; 239 (1-2): 139-47.

16. Bisson I, Prowse DM. WNT signaling regulates self-renewal and differentiation of prostate cancer cells with stem cell characteristics. Cell Res. 2009; 19 (6): 683-97.

17. Bolos V, Blanco M, Medina V, Aparicio G, Diaz-Prado S, Grande E. Notch signalling in cancer stem cells. Clin Transl Oncol. 2009; 11 (1): 11-9.

18. Boman BM, Huang E. Human colon cancer stem cells: a new paradigm in gastrointestinal oncology. J Clin Oncol. 2008; 26 (17): 2828-38.

19. Bonhomme C, Calon A, Martin E, Robine S, Neuville A, Kedinger M, et al. Cdx1, a dispensable homeobox gene for gut development with limited effect in intestinal cancer. Oncogene. 2008; 27 (32): 4497-502.

20. 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.

21. Bos JL, Fearon ER, Hamilton SR, Verlaan-de Vries M, van Boom JH, van der Eb AJ, et al. Prevalence of ras gene mutations in human colorectal cancers. Nature. 1987; 327 (6120): 293-7.

22. Bracht K, Nicholls AM, Liu Y, Bodmer WF. 5-Fluorouracil response in a large panel of colorectal cancer cell lines is associated with mismatch repair deficiency. Br J Cancer. 2010.

23. Bruce WR, Van Der Gaag H. A Quantitative Assay For The Number Of Murine Lymphoma Cells Capable Of Proliferation In Vivo. Nature. 1963; 199: 79-80.

24. Byrd JC, Bresalier RS. Mucins and mucin binding proteins in colorectal cancer. Cancer Metastasis Rev. 2004; 23 (1-2): 77-99.

25. Céspedes MV, Espina C, Garcia-Cabezas MA, Trias M, Boluda A, Gomez del Pulgar MT, et al. Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. Am J Pathol. 2007; 170: 1077-85.

26. Chan CH, Stanners CP. Recent advances in the tumour biology of the GPI-anchored carcinoembryonic antigen family members CEACAM5 and CEACAM6. Curr Oncol. 2007; 14 (2): 70-3.

27. Chan CW, Wong NA, Liu Y, Bicknell D, Turley H, Hollins L, et al. Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. Proc Natl Acad Sci U S A. 2009; 106 (6): 1936-41.

28. Chappell J, Dalton S. Altered cell cycle regulation helps stem-like carcinoma cells resist apoptosis. BMC Biol. 2010; 8: 63.

29. Chen CY, Chiou SH, Huang CY, Jan CI, Lin SC, Tsai ML, et al. Distinct population of highly malignant cells in a head and neck squamous cell carcinoma cell line established by xenograft model. J Biomed Sci. 2009a; 16: 100.

30. Chen YC, Chen YW, Hsu HS, Tseng LM, Huang PI, Lu KH, et al. Aldehyde dehydrogenase 1 is a putative marker for cancer stem cells in head and neck squamous cancer. Biochem Biophys Res Commun. 2009b; 385 (3): 307-13.

31. Chong G, Lee FT, Hopkins W, Tebbutt N, Cebon JS, Mountain AJ, et al. Phase I trial of 131I-huA33 in patients with advanced colorectal carcinoma. Clin Cancer Res. 2005; 11 (13): 4818-26.

32. Chu P, Clanton DJ, Snipas TS, Lee J, Mitchell E, Nguyen ML, et al. Characterization of a subpopulation of colon cancer cells with stem cell-like properties. Int J Cancer. 2009; 124 (6): 1312-21.

33. Cichy J, Pure E. The liberation of CD44. J Cell Biol. 2003; 161 (5): 839-43.

34. 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.

35. Clarke RB. Isolation and characterization of human mammary stem cells. Cell Prolif. 2005; 38 (6): 375-86.

36. Clontech. pZsGreen1-1 Vector Information. Mountain View: Clontech Laboratories; 2004.

37. Clontech. pBI-CMV3 Vector Information. Mountain View: Clontech Laboratories; 2009.

38. CRUK. Latest UK cancer incidence and mortality summary - numbers. [Online] 2009 July [cited 2009 December 28]; Available from: http://info.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@sta/documents/generalcontent/crukmig_1000ast-2735.pdf

39. CRUK. Latest UK cancer incidence and mortality summary - rates. [Online] 2009 July [cited 2009 December 28]; Available from: http://info.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@sta/documents/generalcontent/crukmig_1000ast-2736.pdf

40. Cui H, Ma J, Ding J, Li T, Alam G, Ding HF. Bmi-1 regulates the differentiation and clonogenic self-renewal of I-type neuroblastoma cells in a concentration-dependent manner. J Biol Chem. 2006; 281 (45): 34696-704.

41. Culhane AC, Quackenbush J. Confounding effects in "A six-gene signature predicting breast cancer lung metastasis". Cancer Res. 2009; 69 (18): 7480-5.

42. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A. 2007; 104 (24): 10158-63.

43. Dallas NA, Xia L, Fan F, Gray MJ, Gaur P, van Buren G, 2nd, et al. Chemoresistant colorectal cancer cells, the cancer stem cell phenotype, and increased sensitivity to insulin-like growth factor-I receptor inhibition. Cancer Res. 2009; 69 (5): 1951-7.

44. Denys H, Derycke L, Hendrix A, Westbroek W, Gheldof A, Narine K, et al. Differential impact of TGF-beta and EGF on fibroblast differentiation and invasion reciprocally promotes colon cancer cell invasion. Cancer Lett. 2008; 266 (2): 263-74.

45. Di Fiore R, Santulli A, Ferrante RD, Giuliano M, De Blasio A, Messina C, et al. Identification and expansion of human osteosarcoma-cancer-stem cells by long-term 3-aminobenzamide treatment. J Cell Physiol. 2009; 219 (2): 301-13.

46. Dikomitou O. Genetic and epigenetic characterisation of a panel of CRC cell lines [Doctor of Philosophy thesis]. Oxford: University of Oxford; 2009.

47. Douglas EJ, Fiegler H, Rowan A, Halford S, Bicknell DC, Bodmer W, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. Cancer Res. 2004; 64 (14): 4817-25.

48. Du J, Li Y, Li J, Zheng J. Polycomb group protein Bmi1 expression in colon cancers predicts the survival. Med Oncol. 2009.

49. Du L, Wang H, He L, Zhang J, Ni B, Wang X, et al. CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res. 2008; 14 (21): 6751-60.

50. Durrant LG, Chapman MA, Buckley DJ, Spendlove I, Robins RA, Armitage NC. Enhanced expression of the complement regulatory protein CD55 predicts a poor prognosis in colorectal cancer patients. Cancer Immunol Immunother. 2003; 52 (10): 638-42.

51. Dylla SJ, Beviglia L, Park IK, Chartier C, Raval J, Ngan L, et al. Colorectal cancer stem cells are enriched in xenogeneic tumors following chemotherapy. PLoS One. 2008; 3 (6): e2428.

52. Elsaba TM, Martinez-Pomares L, Robins AR, Crook S, Seth R, Jackson D, et al. The stem cell marker CD133 associates with enhanced colony formation and cell motility in colorectal cancer. PLoS One. 2010; 5 (5): e10714.

53. Espada J, Calvo MB, Diaz-Prado S, Medina V. Wnt signalling and cancer stem cells. Clin Transl Oncol. 2009; 11 (7): 411-27.

54. Fang DD, Kim YJ, Lee CN, Aggarwal S, McKinnon K, Mesmer D, et al. Expansion of CD133(+) colon cancer cultures retaining stem cell properties to enable cancer stem cell target discovery. Br J Cancer. 2010; 102 (8): 1265-75.

55. Feng HL, Liu YQ, Yang LJ, Bian XC, Yang ZL, Gu B, et al. Expression of CD133 correlates with differentiation of human colon cancer cells. Cancer Biol Ther. 2010; 9 (3): 216-23.

56. Fevr T, Robine S, Louvard D, Huelsken J. Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. Mol Cell Biol. 2007; 27 (21): 7551-9.

57. Fodde R, Smits R. Disease model: familial adenomatous polyposis. Trends Mol Med. 2001; 7 (8): 369-73.

58. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. Nat Rev Cancer. 2001; 1 (1): 55-67.

59. Gancz D, Fishelson Z. Cancer resistance to complement-dependent cytotoxicity (CDC): Problem-oriented research and development. Mol Immunol. 2009; 46 (14): 2794-800.

60. Geibel JP. Secretion and absorption by colonic crypts. Annu Rev Physiol. 2005; 67: 471-90.

61. Geiger TR, Peeper DS. Metastasis mechanisms. Biochim Biophys Acta. 2009; 1796 (2): 293-308.

62. Gil J, Bernard D, Peters G. Role of polycomb group proteins in stem cell self-renewal and cancer. DNA Cell Biol. 2005; 24 (2): 117-25.

63. 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.

64. Greaves LC, Preston SL, Tadrous PJ, Taylor RW, Barron MJ, Oukrif D, et al. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. Proc Natl Acad Sci U S A. 2006; 103 (3): 714-9.

65. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology. 2005; 129 (2): 626-38.

66. Guo RJ, Suh ER, Lynch JP. The role of Cdx proteins in intestinal development and cancer. Cancer Biol Ther. 2004; 3 (7): 593-601.

67. Hamburger AW, Salmon SE. Primary bioassay of human tumor stem cells. Science. 1977; 197 (4302): 461-3.

68. Han J, Chang H, Giricz O, Lee GY, Baehner FL, Gray JW, et al. Molecular predictors of 3D morphogenesis by breast cancer cell lines in 3D culture. PLoS Comput Biol. 2010; 6 (2): e1000684.

69. Härma V, Virtanen J, Makela R, Happonen A, Mpindi JP, Knuuttila M, et al. A comprehensive panel of three-dimensional models for studies of prostate cancer growth, invasion and drug responses. PLoS One. 2010; 5 (5): e10431.

70. Harper LJ, Costea DE, Gammon L, Fazil B, Biddle A, Mackenzie IC. Normal and malignant epithelial cells with stem-like properties have an extended G2 cell cycle phase that is associated with apoptotic resistance. BMC Cancer. 2010; 10: 166.

71. He X, Marchionni L, Hansel DE, Yu W, Sood A, Yang J, et al. Differentiation of a highly tumorigenic basal cell compartment in urothelial carcinoma. Stem Cells. 2009; 27 (7): 1487-95.

72. Heath JK, White SJ, Johnstone CN, Catimel B, Simpson RJ, Moritz RL, et al. The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. Proc Natl Acad Sci U S A. 1997; 94 (2): 469-74.

73. Hemenway CS, Halligan BW, Levy LS. The Bmi-1 oncoprotein interacts with dinG and MPh2: the role of RING finger domains. Oncogene. 1998; 16 (19): 2541-7.

74. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell. 2007; 1 (3): 313-23.

75. Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, Gum ET, et al. Heterogeneity of mucin gene expression in normal and neoplastic tissues. Cancer Res. 1993; 53 (3): 641-51.

76. Hoey T, Yen WC, Axelrod F, Basi J, Donigian L, Dylla S, et al. DLL4 blockade inhibits tumor growth and reduces tumor-initiating cell frequency. Cell Stem Cell. 2009; 5 (2): 168-77.

77. Hoffman EM. Inhibition of complement by a substance isolated from human erythrocytes. I. Extraction from human erythrocyte stromata. Immunochemistry. 1969; 6 (3): 391-403.

78. Hopkins DR, Keles S, Greenspan DS. The bone morphogenetic protein 1/Tolloid-like metalloproteinases. Matrix Biol. 2007; 26 (7): 508-23.

79. Horst D, Scheel SK, Liebmann S, Neumann J, Maatz S, Kirchner T, et al. The cancer stem cell marker CD133 has high prognostic impact but unknown functional relevance for the metastasis of human colon cancer. J Pathol. 2009; 219 (4): 427-34.

80. Huang EH, Hynes MJ, Zhang T, Ginestier C, Dontu G, Appelman H, et al. Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. Cancer Res. 2009; 69 (8): 3382-9.

81. Huang M, Li Y, Zhang H, Nan F. Breast cancer stromal fibroblasts promote the generation of CD44+CD24- cells through SDF-1/CXCR4 interaction. J Exp Clin Cancer Res. 2010; 29 (1): 80.

82. Huang P WC, Gou SM, Wu HS, Liu T, Xiong JX. Isolation and biological analysis of tumor stem cells from pancreatic adenocarcinoma. World J Gastroenterol. 2008; 14 (24): 3903-3907.

83. Humphries A, Wright NA. Colonic crypt organization and tumorigenesis. Nat Rev Cancer. 2008; 8 (6): 415-24.

84. Ieta K, Tanaka F, Haraguchi N, Kita Y, Sakashita H, Mimori K, et al. Biological and genetic characteristics of tumor-initiating cells in colon cancer. Ann Surg Oncol. 2008; 15 (2): 638-48.

85. Ikeda J, Morii E, Liu Y, Qiu Y, Nakamichi N, Jokoji R, et al. Prognostic significance of CD55 expression in breast cancer. Clin Cancer Res. 2008; 14 (15): 4780-6.

86. Ilantzis C, DeMarte L, Screaton RA, Stanners CP. Deregulated expression of the human tumor marker CEA and CEA family member CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation. Neoplasia. 2002; 4 (2): 151-63.

87. Invitrogen. AccuPrime(tm) Pfx DNA Polymerase. Carlsbad: Invitrogen Corporation; 2010.

88. Ishimoto T, Oshima H, Oshima M, Kai K, Torii R, Masuko T, et al. CD44+ slow-cycling tumor cell expansion is triggered by cooperative actions of Wnt and prostaglandin E2 in gastric tumorigenesis. Cancer Sci. 2010; 101 (3): 673-8.

89. Jacobs JJ, Kieboom K, Marino S, DePinho RA, van Lohuizen M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature. 1999; 397 (6715): 164-8.

90. Jamieson CH, Ailles LE, Dylla SJ, Muijtjens M, Jones C, Zehnder JL, et al. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. N Engl J Med. 2004; 351 (7): 657-67.

91. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. CA Cancer J Clin. 2010. In press.

92. Jiang F, Qiu Q, Khanna A, Todd NW, Deepak J, Xing L, et al. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. Mol Cancer Res. 2009; 7 (3): 330-8.

93. Johnston MD, Maini PK, Chapman SJ, Edwards CM, Bodmer WF. On the proportion of cancer stem cells in a tumour. J Theor Biol. 2010. In press.

94. Jou T-S, Stewart DB, Stappert J, Nelson WJ, Marrs JA. Genetic and biochemical dissection of protein linkages in the cadherin-catenin complex. Proc Natl Acad Sci U S A. 1995; 92: 5067-71.

95. Kai K, Nagano O, Sugihara E, Arima Y, Sampetrean O, Ishimoto T, et al. Maintenance of HCT116 colon cancer cell line conforms to a stochastic model but not a cancer stem cell model. Cancer Sci. 2009; 100 (12): 2275-82.

96. Kalluri R, Zeisberg M. Fibroblasts in cancer. Nat Rev Cancer. 2006; 6 (5): 392-401.

97. Kawasaki BT, Mistree T, Hurt EM, Kalathur M, Farrar WL. Co-expression of the toleragenic glycoprotein, CD200, with markers for cancer stem cells. Biochem Biophys Res Commun. 2007; 364 (4): 778-82.

98. Keller MS, Ezaki T, Guo RJ, Lynch JP. Cdx1 or Cdx2 expression activates E-cadherin-mediated cell-cell adhesion and compaction in human COLO 205 cells. Am J Physiol Gastrointest Liver Physiol. 2004; 287 (1): G104-14.

99. Kelly PN, Dakic A, Adams JM, Nutt SL, Strasser A. Tumor growth need not be driven by rare cancer stem cells. Science. 2007; 317 (5836): 337.

100. Kemper K, Sprick MR, de Bree M, Scopelliti A, Vermeulen L, Hoek M, et al. The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. Cancer Res. 2010; 70 (2): 719-29.

101. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol Oncol. 2007; 1 (1): 84-96.

102. Kim JH, Yoon SY, Kim CN, Joo JH, Moon SK, Choe IS, et al. The Bmi-1 oncoprotein is overexpressed in human colorectal cancer and correlates with the reduced p16INK4a/p14ARF proteins. Cancer Lett. 2004; 203 (2): 217-24.

103. Kim KM, Shibata D. Methylation reveals a niche: stem cell succession in human colon crypts. Oncogene. 2002; 21 (35): 5441-9.

104. Kim MP, Evans DB, Wang H, Abbruzzese JL, Fleming JB, Gallick GE. Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. Nat Protoc. 2009; 4 (11): 1670-80.

105. Kinsella AR, Lepts GC, Hill CL, Jones M. Reduced E-cadherin expression correlates with increased invasiveness in colorectal carcinoma cell lines. Clin Exp Metastasis. 1994; 12 (4): 335-42.

106. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996; 87 (2): 159-70.

107. Kirkland SC. Type I collagen inhibits differentiation and promotes a stem cell-like phenotype in human colorectal carcinoma cells. Br J Cancer. 2009; 101 (2): 320-6.

108. Kirkland SC, Ying H. Alpha2beta1 integrin regulates lineage commitment in multipotent human colorectal cancer cells. J Biol Chem. 2008; 283 (41): 27612-9.

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

110. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971; 68 (4): 820-3.

111. Kontos CK, Papadopoulos IN, Scorilas A. Quantitative expression analysis and prognostic significance of the novel apoptosis-related gene BCL2L12 in colon cancer. Biol Chem. 2008; 389 (12): 1467-75.

112. Koretz K, Bruderlein S, Henne C, Moller P. Decay-accelerating factor (DAF, CD55) in normal colorectal mucosa, adenomas and carcinomas. Br J Cancer. 1992; 66 (5): 810-4.

113. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science. 1997; 275 (5307): 1784-7.

114. Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature. 2006; 442 (7104): 818-22.

115. Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. Curr Opin Cell Biol. 2006; 18 (5): 565-71.

116. Lang SH, Sharrard RM, Stark M, Villette JM, Maitland NJ. Prostate epithelial cell lines form spheroids with evidence of glandular differentiation in three-dimensional Matrigel cultures. Br J Cancer. 2001; 85 (4): 590-9.

117. Lawrie LC, Dundas SR, Curran S, Murray GI. Liver fatty acid binding protein expression in colorectal neoplasia. Br J Cancer. 2004; 90 (10): 1955-60.

118. Le NH, Franken P, Fodde R. Tumour-stroma interactions in colorectal cancer: converging on beta-catenin activation and cancer stemness. Br J Cancer. 2008; 98 (12): 1886-93.

119. Lee CJ, Dosch J, Simeone DM. Pancreatic cancer stem cells. J Clin Oncol. 2008; 26 (17): 2806-12.

120. Lee S, Bang S, Song K, Lee I. Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon. Oncol Rep. 2006; 16 (4): 747-54.

121. Levings PP, McGarry SV, Currie TP, Nickerson DM, McClellan S, Ghivizzani SC, et al. Expression of an exogenous human Oct-4 promoter identifies tumor-initiating cells in osteosarcoma. Cancer Res. 2009; 69 (14): 5648-55.

122. Li C, Lee CJ, Simeone DM. Identification of human pancreatic cancer stem cells. Methods Mol Biol. 2009; 568: 161-73.

123. Li DW, Tang HM, Fan JW, Yan DW, Zhou CZ, Li SX, et al. Expression level of Bmi-1 oncoprotein is associated with progression and prognosis in colon cancer. J Cancer Res Clin Oncol. 2010; 136 (7): 997-1006.

124. Li L, Spendlove I, Morgan J, Durrant LG. CD55 is over-expressed in the tumour environment. Br J Cancer. 2001; 84 (1): 80-6.

125. Liao D, Luo Y, Markowitz D, Xiang R, Reisfeld RA. Cancer associated fibroblasts promote tumor growth and metastasis by modulating the tumor immune microenvironment in a 4T1 murine breast cancer model. PLoS One. 2009; 4 (11): e7965.

126. Lim SC. CD24 and human carcinoma: tumor biological aspects. Biomed Pharmacother. 2005; 59 Suppl 2: S351-4.

127. Liu L, Andrews LG, Tollefsbol TO. Loss of the human polycomb group protein BMI1 promotes cancer-specific cell death. Oncogene. 2006a; 25 (31): 4370-5.

128. Liu S, Dontu G, Mantle ID, Patel S, Ahn NS, Jackson KW, et al. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. Cancer Res. 2006b; 66 (12): 6063-71.

129. Loberg RD, Day LL, Dunn R, Kalikin LM, Pienta KJ. Inhibition of decay-accelerating factor (CD55) attenuates prostate cancer growth and survival in vivo. Neoplasia. 2006; 8 (1): 69-78.

130. Lobo NA, Shimono Y, Qian D, Clarke MF. The biology of cancer stem cells. Annu Rev Cell Dev Biol. 2007; 23: 675-99.

131. Lugli A, Iezzi G, Hostettler I, Muraro MG, Mele V, Tornillo L, et al. Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. Br J Cancer. 2010.

132. Lynch J, Keller M, Guo RJ, Yang D, Traber P. Cdx1 inhibits the proliferation of human colon cancer cells by reducing cyclin D1 gene expression. Oncogene. 2003; 22 (41): 6395-407.

133. Mackenzie IC. Growth of malignant oral epithelial stem cells after seeding into organotypical cultures of normal mucosa. J Oral Pathol Med. 2004; 33 (2): 71-8.

134. Madison BB, Braunstein K, Kuizon E, Portman K, Qiao XT, Gumucio DL. Epithelial hedgehog signals pattern the intestinal crypt-villus axis. Development. 2005; 132 (2): 279-89.

135. Maeda M, Johnson KR, Wheelock MJ. Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition. J Cell Sci. 2005; 118 (Pt 5): 873-87.

136. Makrilia N, Kollias A, Manolopoulos L, Syrigos K. Cell adhesion molecules: role and clinical significance in cancer. Cancer Invest. 2009; 27 (10): 1023-37.

137. Mallo GV, Soubeyran P, Lissitzky JC, Andre F, Farnarier C, Marvaldi J, et al. Expression of the Cdx1 and Cdx2 homeotic genes leads to reduced malignancy in colon cancer-derived cells. J Biol Chem. 1998; 273 (22): 14030-6.

138. Mao L, Xia YP, Zhou YN, Dai RL, Yang X, Duan SJ, et al. A critical role of Sonic Hedgehog signaling in maintaining the tumorigenicity of neuroblastoma cells. Cancer Sci. 2009; 100 (10): 1848-55.

139. Marx J. Cancer biology. All in the stroma: cancer's Cosa Nostra. Science. 2008; 320 (5872): 38-41.

140. McKaig BC, Makh SS, Hawkey CJ, Podolsky DK, Mahida YR. Normal human colonic subepithelial myofibroblasts enhance epithelial migration (restitution) via TGF-beta3. Am J Physiol. 1999; 276 (5 Pt 1): G1087-93.

141. McLaughlin PM, Trzpis M, Kroesen BJ, Helfrich W, Terpstra P, Dokter WH, et al. Use of the EGP-2/Ep-CAM promoter for targeted expression of heterologous genes in carcinoma derived cell lines. Cancer Gene Ther. 2004; 11 (9): 603-12.

142. Medina V, Calvo MB, Diaz-Prado S, Espada J. Hedgehog signalling as a target in cancer stem cells. Clin Transl Oncol. 2009; 11 (4): 199-207.

143. Mikesch JH, Buerger H, Simon R, Brandt B. Decay-accelerating factor (CD55): a versatile acting molecule in human malignancies. Biochim Biophys Acta. 2006; 1766 (1): 42-52.

144. Miyaki M, Konishi M, Kikuchi-Yanoshita R, Enomoto M, Igari T, Tanaka K, et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res. 1994; 54 (11): 3011-20.

145. Molofsky AV, Pardal R, Iwashita T, Park IK, Clarke MF, Morrison SJ. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature. 2003; 425 (6961): 962-7.

146. Mourra N, Zeitoun G, Portier G, Blanche H, Tubacher E, Gressin L, et al. High-resolution genotyping of chromosome 8 in colon adenocarcinomas reveals recurrent break point but no gene mutation in the 8p21 region. Diagn Mol Pathol. 2008; 17: 90-3.

147. Mueller L, Goumas FA, Affeldt M, Sandtner S, Gehling UM, Brilloff S, et al. Stromal fibroblasts in colorectal liver metastases originate from resident fibroblasts and generate an inflammatory microenvironment. Am J Pathol. 2007; 171 (5): 1608-18.

148. Nakai E, Park K, Yawata T, Chihara T, Kumazawa A, Nakabayashi H, et al. Enhanced MDR1 expression and chemoresistance of cancer stem cells derived from glioblastoma. Cancer Invest. 2009; 27 (9): 901-8.

149. Navarro-Alvarez N, Kondo E, Kawamoto H, Hassan W, Yuasa T, Kubota Y, et al. Isolation and Propagation of a Human CD133-Negative Colon Tumor Derived Cell Line with Tumorigenic and Angiogenic Properties. Cell Transplant. 2010. In press

150. Navarro-Alvarez N, Soto-Gutierrez A, Yuasa T, Yamatsuji T, Shirakawa Y, Nagasaka T, et al. Long-term culture of Japanese human embryonic stem cells in feeder-free conditions. Cell Transplant. 2008; 17 (1-2): 27-33.

151. Nishimura S, Wakabayashi N, Toyoda K, Kashima K, Mitsufuji S. Expression of Musashi-1 in human normal colon crypt cells: a possible stem cell marker of human colon epithelium. Dig Dis Sci. 2003; 48 (8): 1523-9.

152. Novelli M, Cossu A, Oukrif D, Quaglia A, Lakhani S, Poulsom R, et al. X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. Proc Natl Acad Sci U S A. 2003; 100 (6): 3311-4.

153. Novelli MR, Williamson JA, Tomlinson IP, Elia G, Hodgson SV, Talbot IC, et al. Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. Science. 1996; 272 (5265): 1187-90.

154. 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.

155. Odening KE, Li W, Rutz R, Laufs S, Fruehauf S, Fishelson Z, et al. Enhanced complement resistance in drug-selected P-glycoprotein expressing multi-drug-resistant ovarian carcinoma cells. Clin Exp Immunol. 2009; 155 (2): 239-48.

156. Ohtsuki S, Kamoi M, Watanabe Y, Suzuki H, Hori S, Terasaki T. Correlation of induction of ATP binding cassette transporter A5 (ABCA5) and ABCB1 mRNAs with differentiation state of human colon tumor. Biol Pharm Bull. 2007; 30 (6): 1144-6.

157. Ootani A, Li X, Sangiorgi E, Ho QT, Ueno H, Toda S, et al. Sustained in vitro intestinal epithelial culture within a Wnt-dependent stem cell niche. Nat Med. 2009; 15 (6): 701-6.

158. Ostman A, Augsten M. Cancer-associated fibroblasts and tumor growth--bystanders turning into key players. Curr Opin Genet Dev. 2009; 19 (1): 67-73.

159. Pacheco, II, Macleod RJ. CaSR stimulates secretion of Wnt5a from colonic myofibroblasts to stimulate CDX2 and sucrase-isomaltase using Ror2 on intestinal epithelia. Am J Physiol Gastrointest Liver Physiol. 2008; 295 (4): G748-59.

160. Pang R, Law WL, Chu AC, Poon JT, Lam CS, Chow AK, et al. A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. Cell Stem Cell. 2010; 6 (6): 603-15.

161. Phillips TM, McBride WH, Pajonk F. The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation. J Natl Cancer Inst. 2006; 98 (24): 1777-85.

162. Pignatelli M, Bodmer WF. Genetics and biochemistry of collagen binding-triggered glandular differentiation in a human colon carcinoma cell line. Proc Natl Acad Sci U S A. 1988; 85 (15): 5561-5.

163. Pignatelli M, Bodmer WF. Integrin-receptor-mediated differentiation and growth inhibition are enhanced by transforming growth factor-beta in colorectal tumour cells grown in collagen gel. Int J Cancer. 1989; 44 (3): 518-23.

164. Pignatelli M, Smith ME, Bodmer WF. Low expression of collagen receptors in moderate and poorly differentiated colorectal adenocarcinomas. Br J Cancer. 1990; 61 (4): 636-8.

165. Pilozzi E, Onelli MR, Ziparo V, Mercantini P, Ruco L. CDX1 expression is reduced in colorectal carcinoma and is associated with promoter hypermethylation. J Pathol. 2004; 204 (3): 289-95.

166. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol. 2003; 4 (1): 33-45.

167. Powell DW, Mifflin RC, Valentich JD, Crowe SE, Saada JI, West AB. Myofibroblasts. I. Paracrine cells important in health and disease. Am J Physiol. 1999a; 277 (1 Pt 1): C1-9.

168. Powell DW, Mifflin RC, Valentich JD, Crowe SE, Saada JI, West AB. Myofibroblasts. II. Intestinal subepithelial myofibroblasts. Am J Physiol. 1999b; 277 (2 Pt 1): C183-201.

169. Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, et al. APC mutations occur early during colorectal tumorigenesis. Nature. 1992; 359 (6392): 235-7.

170. Pratap J, Lian JB, Javed A, Barnes GL, van Wijnen AJ, Stein JL, et al. Regulatory roles of Runx2 in metastatic tumor and cancer cell interactions with bone. Cancer Metastasis Rev. 2006; 25 (4): 589-600.

171. Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proc Natl Acad Sci U S A. 2007; 104 (3): 973-8.

172. Promega. CellTiter 96(r) Aqueous Non-Radioactive Cell Proliferation Assay. Madison: Promega Corporation; 2009.

173. Quintana E, Shackleton M, Sabel MS, Fullen DR, Johnson TM, Morrison SJ. Efficient tumour formation by single human melanoma cells. Nature. 2008; 456 (7222): 593-8.

174. Rageul J, Mottier S, Jarry A, Shah Y, Theoleyre S, Masson D, et al. KLF4-dependent, PPARgamma-induced expression of GPA33 in colon cancer cell lines. Int J Cancer. 2009; 125 (12): 2802-9.

175. Rathinam R, Alahari SK. Important role of integrins in the cancer biology. Cancer Metastasis Rev. 2010; 29 (1): 223-37.

176. Renkonen ET, Nieminen P, Abdel-Rahman WM, Moisio A-L, Jarvela I, Arte S, et al. Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J Clin Oncol. 2005; 23: 5651-9.

177. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. Nature. 2007; 445 (7123): 111-5.

178. Richman PI, Bodmer WF. Control of differentiation in human colorectal carcinoma cell lines: epithelial-mesenchymal interactions. J Pathol. 1988; 156 (3): 197-211.

179. Richman PI, Tilly R, Jass JR, Bodmer WF. Colonic pericrypt sheath cells: characterisation of cell type with new monoclonal antibody. J Clin Pathol. 1987; 40 (6): 593-600.

180. Roig AI, Eskiocak U, Hight SK, Kim SB, Delgado O, Souza RF, et al. Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. Gastroenterology. 2010; 138 (3): 1012-21 e1-5.

181. Roth U, Razawi H, Hommer J, Engelmann K, Schwientek T, Muller S, et al. Differential expression proteomics of human colorectal cancer based on a syngeneic cellular model for the progression of adenoma to carcinoma. Proteomics. 2010; 10 (2): 194-202.

182. Rutzky LP, Kaye CI, Siciliano MJ, Chao M, Kahan BD. Longitudinal karyotype and genetic signature analysis of cultured human colon adenocarcinoma cell lines LS180 and LS174T. Cancer Res. 1980; 40 (5): 1443-8.

183. Samoszuk M, Tan J, Chorn G. Clonogenic growth of human breast cancer cells co-cultured in direct contact with serum-activated fibroblasts. Breast Cancer Res. 2005; 7 (3): R274-83.

184. Sangiorgi E, Capecchi MR. Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet. 2008; 40 (7): 915-20.

185. Santisteban M, Reiman JM, Asiedu MK, Behrens MD, Nassar A, Kalli KR, et al. Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. Cancer Res. 2009; 69 (7): 2887-95.

186. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009; 459 (7244): 262-5.

187. Schatton T, Murphy GF, Frank NY, Yamaura K, Waaga-Gasser AM, Gasser M, et al. Identification of cells initiating human melanomas. Nature. 2008; 451 (7176): 345-9.

188. Schulenburg A, Cech P, Herbacek I, Marian B, Wrba F, Valent P, et al. CD44-positive colorectal adenoma cells express the potential stem cell markers musashi antigen (msi1) and ephrin B2 receptor (EphB2). J Pathol. 2007; 213 (2): 152-60.

189. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009; 138 (5): 822-9.

190. Shimoda M, Mellody KT, Orimo A. Carcinoma-associated fibroblasts are a rate-limiting determinant for tumour progression. Semin Cell Dev Biol. 2010; 21 (1): 19-25.

191. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell. 2007; 11 (3): 259-73.

192. Shmelkov SV, Butler JM, Hooper AT, Hormigo A, Kushner J, Milde T, et al. CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. J Clin Invest. 2008; 118 (6): 2111-20.

193. Silberg DG, Furth EE, Taylor JK, Schuck T, Chiou T, Traber PG. CDX1 protein expression in normal, metaplastic, and neoplastic human alimentary tract epithelium. Gastroenterology. 1997; 113 (2): 478-86.

194. Simon JA, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol. 2009; 10 (10): 697-708.

195. 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.

196. Slevin M, Payne S. New treatments for colon cancer. BMJ. 2004; 329 (7458): 124-6.

197. Su Y, Qiu Q, Zhang X, Jiang Z, Leng Q, Liu Z, et al. Aldehyde dehydrogenase 1 A1-positive cell population is enriched in tumor-initiating cells and associated with progression of bladder cancer. Cancer Epidemiol Biomarkers Prev. 2010; 19 (2): 327-37.

198. Subramaniam V, Vincent IR, Gilakjan M, Jothy S. Suppression of human colon cancer tumors in nude mice by siRNA CD44 gene therapy. Exp Mol Pathol. 2007; 83 (3): 332-40.

199. Takaishi S, Okumura T, Tu S, Wang SS, Shibata W, Vigneshwaran R, et al. Identification of gastric cancer stem cells using the cell surface marker CD44. Stem Cells. 2009; 27 (5): 1006-20.

200. Tamase A, Muraguchi T, Naka K, Tanaka S, Kinoshita M, Hoshii T, et al. Identification of tumor-initiating cells in a highly aggressive brain tumor using promoter activity of nucleostemin. Proc Natl Acad Sci U S A. 2009; 106 (40): 17163-8.

201. Tan P, Lee T. Identification of ALDH-expressing cancer stem cells. Vancouver: STEMCELL Technologies; 2009.

202. Tateishi K, Ohta M, Kanai F, Guleng B, Tanaka Y, Asaoka Y, et al. Dysregulated expression of stem cell factor Bmi1 in precancerous lesions of the gastrointestinal tract. Clin Cancer Res. 2006; 12 (23): 6960-6.

203. Thawani JP, Wang AC, Than KD, Lin CY, La Marca F, Park P. Bone morphogenetic proteins and cancer: review of the literature. Neurosurgery. 2010; 66 (2): 233-46; discussion 246.

204. Thermo. NanoDrop 1000 Spectrophotometer V3.7 User's Manual. Wilmington: Thermo Scientific; 2008.

205. Till JE, Mc CE. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res. 1961; 14: 213-22.

206. Todaro M, Alea MP, Di Stefano AB, Cammareri P, Vermeulen L, Iovino F, et al. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. Cell Stem Cell. 2007; 1 (4): 389-402.

207. Todaro M, Perez Alea M, Scopelliti A, Medema JP, Stassi G. IL-4-mediated drug resistance in colon cancer stem cells. Cell Cycle. 2008; 7 (3): 309-13.

208. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. Nat Rev Cancer. 2010; 10 (2): 116-29.

209. Uchino H, Kataoka H, Itoh H, Koono M. Expression of intestinal trefoil factor mRNA is downregulated during progression of colorectal carcinomas. J Clin Pathol. 1997; 50 (11): 932-4.

210. Valentich JD, Popov V, Saada JI, Powell DW. Phenotypic characterization of an intestinal subepithelial myofibroblast cell line. Am J Physiol. 1997; 272 (5 Pt 1): C1513-24.

211. Varnat F, Duquet A, Malerba M, Zbinden M, Mas C, Gervaz P, et al. Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. EMBO Mol Med. 2009; 1 (6-7): 338-51.

212. Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T, et al. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol. 2010; 12 (5): 468-76.

213. Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea M, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. Proc Natl Acad Sci U S A. 2008; 105 (36): 13427-32.

214. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med. 1988; 319 (9): 525-32.

215. Volchenboum SL, Li C, Li S, Attiyeh EF, Reynolds CP, Maris JM, et al. Comparison of primary neuroblastoma tumors and derivative early-passage cell lines using genome-wide single nucleotide polymorphism array analysis. Cancer Res. 2009; 69 (10): 4143-9.

216. Wai PY, Mi Z, Gao C, Guo H, Marroquin C, Kuo PC. Ets-1 and runx2 regulate transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells. J Biol Chem. 2006; 281 (28): 18973-82.

217. Wang J, Wakeman TP, Lathia JD, Hjelmeland AB, Wang XF, White RR, et al. Notch promotes radioresistance of glioma stem cells. Stem Cells. 2010; 28 (1): 17-28.

218. Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. Trends Cell Biol. 2005; 15 (9): 494-501.

219. Wang W, Li Q, Yamada T, Matsumoto K, Matsumoto I, Oda M, et al. Crosstalk to stromal fibroblasts induces resistance of lung cancer to epidermal growth factor receptor tyrosine kinase inhibitors. Clin Cancer Res. 2009; 15 (21): 6630-8.

220. Wong NA, Britton MP, Choi GS, Stanton TK, Bicknell DC, Wilding JL, et al. Loss of CDX1 expression in colorectal carcinoma: promoter methylation, mutation, and loss of heterozygosity analyses of 37 cell lines. Proc Natl Acad Sci U S A. 2004; 101 (2): 574-9.

221. Wu WK, Sung JJ, Wu YC, Li ZJ, Yu L, Cho CH. Bone morphogenetic protein signalling is required for the anti-mitogenic effect of the proteasome inhibitor MG-132 on colon cancer cells. Br J Pharmacol. 2008; 154 (3): 632-8.

222. Xi L, Gooding W, McCarty K, Godfrey TE, Hughes SJ. Identification of mRNA markers for molecular staging of lymph nodes in colorectal cancer. Clin Chem. 2006; 52 (3): 520-3.

223. Xu JX, Morii E, Liu Y, Nakamichi N, Ikeda J, Kimura H, et al. High tolerance to apoptotic stimuli induced by serum depletion and ceramide in side-population cells: high expression of CD55 as a novel character for side-population. Exp Cell Res. 2007; 313 (9): 1877-85.

224. Yamamichi N, Inada K, Furukawa C, Sakurai K, Tando T, Ishizaka A, et al. Cdx2 and the Brm-type SWI/SNF complex cooperatively regulate villin expression in gastrointestinal cells. Exp Cell Res. 2009; 315 (10): 1779-89.

225. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology. 2009; 136 (3): 1012-24.

226. Yang SH, Lin JK, Lai CR, Chen CC, Li AF, Liang WY, et al. Risk factors for peritoneal dissemination of colorectal cancer. J Surg Oncol. 2004; 87 (4): 167-73.

227. Yang YM, Chang JW. Bladder cancer initiating cells (BCICs) are among EMA-CD44v6+ subset: novel methods for isolating undetermined cancer stem (initiating) cells. Cancer Invest. 2008; 26 (7): 725-33.

228. Yen TH, Wright NA. The gastrointestinal tract stem cell niche. Stem Cell Rev. 2006; 2 (3): 203-12.

229. Yeung TM, Gandhi SC, Wilding JL, Muschel R, Bodmer WF. Cancer stem cells from colorectal cancer-derived cell lines. Proc Natl Acad Sci U S A. 2010; 107 (8): 3722-7.

230. Yoshikawa K, Shimada M, Miyamoto H, Higashijima J, Miyatani T, Nishioka M, et al. Sonic hedgehog relates to colorectal carcinogenesis. J Gastroenterol. 2009; 44 (11): 1113-7.

231. Zhang S, Balch C, Chan MW, Lai HC, Matei D, Schilder JM, et al. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res. 2008; 68 (11): 4311-20.

232. Zheng X, Shen G, Yang X, Liu W. Most C6 cells are cancer stem cells: evidence from clonal and population analyses. Cancer Res. 2007; 67 (8): 3691-7.

233. Zucchi I, Sanzone S, Astigiano S, Pelucchi P, Scotti M, Valsecchi V, et al. The properties of a mammary gland cancer stem cell. Proc Natl Acad Sci U S A. 2007; 104 (25): 10476-81.

APPENDIX A

CELL LINE CHARACTERIZATION

Table A.1 Cell lines studied and tumor origin characteristics.

Cell Line	Organ Site	Dukes Stage	Histopathological Differentiation
C80	Rectum	D	–
C84	Cecum	C	Poor
C99	Rectum	C	Moderate
CC20	Colon	–	Well
CCK81	–	–	–
Colo741	Pelvic metastasis	D	–
DLD1	Colon	C	–
Gp2D	Colon	B	Poor
HCA7	Hepatic flexure	B	Moderate
HCT116	Colon	–	–
HCT15	Colon	–	–
HRA19	Rectum	B	Well
HT29	Colon	–	Moderate
LIM1863	Cecum	C	Poor
Lovo	Distant lymph node	D	–
LS174T	Colon	B	Well
LS180	Colon	B	Well
LS411	Cecum	B	Poor
LS513	–	–	–
NCIH716	–	–	–
PCJW	Colon	C	Poor
RCM1	–	–	–
RKO	–	–	–
SKCO1	Ascitic fluid	D	–
SNUC2B	–	–	–
SW1222	Colon	C	Moderate
SW1417	Colon	C	–
SW48	Colon	C	Poor
Vaco5	Cecum	C	Poor
WIDR	Recto-sigmoid	–	Moderate

Table A.2 Cell lines studied and known protein mutation and replication error (RER) statuses. Unless otherwise specified, "wt" indicated a wild-type cDNA sequence present in the cell line.

Cell Line	APC	E-cadherin	β-catenin	kRAS	SMAD4	p53	RER
C80	L629X	wt	wt	A146V	D351H	Q52X	–
C84	R283X, R1450X	wt	wt	G12A	wt	R342X	–
C99	wt	wt	wt	wt	wt	wt	–
CC20	wt	wt	wt	wt	wt	Y126X	–
CCK81	Y159C	wt	T41A	wt	wt	wt	+
Colo741	wt	wt	wt	wt	wt	K321X	–
DLD1	I1417X, R2166X	wt	wt	wt	wt	S241F	+
Gp2D	T1445X	P126X	D755V	G12D	wt	wt	+
HCA7	T1556X	wt	wt	wt	wt	R282W	+
HCT116	wt	wt	Δ45S	wt	wt	wt	+
HCT15	I1417X, R2166X	wt	wt	G13D	wt	S241F	+
HRA19	R405X, R1450X	wt	wt	wt	wt	R273H	–
HT29	T1556X, E853X	wt	wt	wt	wt	R273H	–
LIM1863	L1488X	wt	wt	wt	wt	Y234H	–
Lovo	R1114X	wt	wt	G13D	wt	wt	+
LS174T	wt	wt	Δ45S	G12D	wt	wt	–
LS180	R1788C	wt	Δ45S	G12D	wt	wt	+
LS411	Q789X, E1309X	wt	wt	wt	wt	Y126X	+
LS513	wt	wt	wt	G12D	wt	wt	–
NCIH716	wt	wt	wt	wt	wt	E224D	–
PCJW	truncation	wt	wt	wt	wt	wt	–
RCM1	wt	wt	wt	G12V	wt	R306X	–
RKO	L1454E	wt	wt	wt	wt	wt	+
SKCO1	F1089X	wt	wt	G12V	wt	wt	–
SNUC2B	wt	wt	wt	G12D	wt	R273C, R273H	+
SW1222	E1306X	wt	wt	wt	L540R	wt	–
SW1417	R1450X	wt	wt	wt	wt	R72P	–
SW48	R2714C	wt	wt	wt	wt	wt	+
Vaco5	T1556X	wt	wt	wt	wt	R282W	+
WIDR	T1556X	wt	wt	wt	wt	R273H	–

APPENDIX B

RESULTS OF MICROARRAY COMPARISONS

Table B.1 List of genes upregulated in dense lines, sorted by *p*-value. All genes are statistically significant at a significance level of 0.02 and are upregulated at least 1.5-fold in dense lines.

Gene Name	Abbreviation	p-value	Fold-Change
protein kinase C, theta	PRKCQ	0.0000	1.5
plectin 1, intermediate filament binding protein 500kDa	PLEC1	0.0001	1.7
Rho guanine nucleotide exchange factor (GEF) 12	ARHGEF12	0.0002	1.5
runt-related transcription factor 2	RUNX2	0.0002	6.6
Transcribed locus	---	0.0003	1.9
ATPase, class VI, type 11B	ATP11B	0.0003	1.9
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0005	2.1
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0005	2.0
zinc finger CCCH-type containing 12C	ZC3H12C	0.0006	4.1
DnaJ (Hsp40) homolog, subfamily B, member 4	DNAJB4	0.0007	1.8
tripartite motif-containing 23	TRIM23	0.0007	1.9
small nucleolar RNA, C/D box 50A /// small nucleolar RNA host gene (non-protein	SNHG5 /// SNORD50A /// SNORD50B	0.0009	2.3
nucleosome assembly protein 1-like 5	NAP1L5	0.0009	2.6
tripartite motif-containing 23	TRIM23	0.0009	1.7
TNF receptor-associated factor 4	TRAF4	0.0010	1.5
Mediator complex subunit 28	MED28	0.0011	1.7
chromosome 4 open reading frame 15	C4orf15	0.0011	1.7
solute carrier family 23 (nucleobase transporters), member 2	SLC23A2	0.0011	1.5
CDNA FLJ37584 fis, clone BRCOC2004950	---	0.0011	1.5
clusterin	CLU	0.0013	3.8
bone morphogenetic protein 1	BMP1	0.0013	1.6
U2AF homology motif (UHM) kinase 1	UHMK1	0.0013	1.6
family with sequence similarity 63, member B	FAM63B	0.0014	1.7
nuclear receptor subfamily 1, group D, member 2	NR1D2	0.0014	1.7
zinc finger, DHHC-type containing 2	ZDHHC2	0.0015	3.8
neurofibromin 2 (merlin)	NF2	0.0015	1.7
RUN and FYVE domain containing 2	RUFY2	0.0016	1.5
bone morphogenetic protein 1	BMP1	0.0016	1.6
karyopherin alpha 5 (importin alpha 6)	KPNA5	0.0016	2.0
BMI1 polycomb ring finger oncogene	BMI1	0.0019	1.5
TAO kinase 2	TAOK2	0.0020	1.5
Rho GTPase activating protein 29	ARHGAP29	0.0021	1.5
stress-associated endoplasmic reticulum protein family member 2	SERP2	0.0022	2.5
KIAA1946	KIAA1946	0.0024	6.8
pecanex homolog (Drosophila)	PCNX	0.0024	1.5
GTP-binding protein 8 (putative)	GTPBP8	0.0024	1.5
clusterin	CLU	0.0024	5.0
host cell factor C2	HCFC2	0.0024	1.6
ubiquitin specific peptidase 11	USP11	0.0025	1.7

transmembrane protein 136	TMEM136	0.0025	1.5
serine/threonine kinase 17a	STK17A	0.0026	2.8
Stress-associated endoplasmic reticulum protein family member 2	SERP2	0.0026	1.5
interleukin 6 signal transducer (gp130, oncostatin M receptor)	IL6ST	0.0026	2.3
nucleosome assembly protein 1-like 5	NAP1L5	0.0026	3.5
Bardet-Biedl syndrome 7	BBS7	0.0026	2.0
chimerin (chimaerin) 1	CHN1	0.0026	2.7
YOD1 OTU deubiquinating enzyme 1 homolog (S. cerevisiae)	YOD1	0.0026	1.8
DnaJ (Hsp40) homolog, subfamily B, member 14	DNAJB14	0.0026	1.5
sphingomyelin synthase 2	SGMS2	0.0027	2.0
sterile alpha motif domain containing 8	SAMD8	0.0027	1.6
CDNA FLJ34038 fis, clone FCBBF2005645	---	0.0027	2.3
suppressor of cytokine signaling 5	SOCS5	0.0029	1.7
polypyrimidine tract binding protein 2	PTBP2	0.0030	1.5
scavenger receptor class A, member 3	SCARA3	0.0031	2.7
TEA domain family member 1 (SV40 transcriptional enhancer factor)	TEAD1	0.0031	1.7
stearoyl-CoA desaturase 5	SCD5	0.0031	1.8
trafficking protein particle complex 6B	TRAPPC6B	0.0032	1.5
glypican 1	GPC1	0.0033	2.0
PTPRF interacting protein, binding protein 1 (liprin beta 1)	PPFIBP1	0.0034	1.8
WW domain containing transcription regulator 1	WWTR1	0.0034	1.9
adaptor-related protein complex 1, gamma 1 subunit	AP1G1	0.0035	1.5
HECT domain containing 2	HECTD2	0.0035	1.8
hypothetical protein LOC286109	LOC286109	0.0035	1.5
fibroblast growth factor 2 (basic)	FGF2	0.0036	2.1
caveolin 1, caveolae protein, 22kDa	CAV1	0.0037	5.7
ATPase, class VI, type 11B	ATP11B	0.0037	1.6
transforming growth factor, alpha	TGFA	0.0038	1.7
natural killer-tumor recognition sequence	NKTR	0.0038	1.5
purine-rich element binding protein A	PURA	0.0038	1.7
F-box and WD repeat domain containing 2	FBXW2	0.0039	1.5
calmodulin regulated spectrin-associated protein 1-like 1	CAMSAP1L1	0.0040	2.0
hypothetical protein LOC338620	LOC338620	0.0040	1.6
piggyBac transposable element derived 1	PGBD1	0.0041	1.8
UDP-glucose ceramide glucosyltransferase	UGCG	0.0041	1.7
zinc finger protein 131	ZNF131	0.0042	1.8
ribosomal protein S27a	RPS27A	0.0043	1.5
SAR1 gene homolog A (S. cerevisiae)	SAR1A	0.0043	1.7
t-complex 11 (mouse)-like 1	TCP11L1	0.0043	1.8
Chromosome 1 open reading frame 58	C1orf58	0.0044	1.5
tripartite motif-containing 23	TRIM23	0.0044	1.7
four jointed box 1 (Drosophila)	FJX1	0.0044	3.0
neurofibromin 2 (merlin)	NF2	0.0044	2.0

fibroblast growth factor 2 (basic)	FGF2	0.0044	2.5
Mof4 family associated protein 1	MRFAP1	0.0045	1.5
zinc finger and BTB domain containing 47	ZBTB47	0.0045	1.8
caveolin 1, caveolae protein, 22kDa	CAV1	0.0045	4.7
ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	UBE2D3	0.0046	1.5
integrin, beta 8	ITGB8	0.0046	1.5
septin 3	SEP3	0.0046	1.8
suppressor of cytokine signaling 5	SOCS5	0.0046	1.8
myosin phosphatase-Rho interacting protein	M-RIP	0.0046	1.5
basic transcription factor 3-like 4	BTF3L4	0.0047	1.8
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0047	2.3
DCN1, defective in cullin neddylation 1, domain containing 1 (S. cerevisiae)	DCUN1D1	0.0048	1.7
RUN and FYVE domain containing 2	RUFY2	0.0049	1.6
chromosome 8 open reading frame 58	C8orf58	0.0049	1.6
dipeptidyl-peptidase 8	DPP8	0.0049	1.6
G protein-coupled receptor 110	GPR110	0.0050	3.4
EH-domain containing 1	EHD1	0.0050	1.9
SWAP-70 protein	SWAP70	0.0050	2.1
leucine rich repeat containing 8 family, member C	LRRC8C	0.0051	1.7
FERM domain containing 6	FRMD6	0.0051	3.5
TEA domain family member 1 (SV40 transcriptional enhancer factor)	TEAD1	0.0052	1.6
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0052	1.5
PDZ and LIM domain 7 (enigma)	PDLIM7	0.0052	1.6
tubulin folding cofactor E-like	TBCEL	0.0053	1.9
ubiquitin specific peptidase 38	USP38	0.0053	1.6
four and a half LIM domains 3	FHL3	0.0053	1.5
chromosome 10 open reading frame 12	C10orf12	0.0053	1.5
SWAP-70 protein	SWAP70	0.0054	2.3
TROVE domain family, member 2	TROVE2	0.0054	1.8
WW and C2 domain containing 2	WWC2	0.0054	2.3
SERTA domain containing 1	SERTAD1	0.0055	1.7
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0055	1.7
SH3 domain containing ring finger 1	SH3RF1	0.0056	1.7
FERM domain containing 6	FRMD6	0.0056	3.0
chromosome 22 open reading frame 29	C22orf29	0.0056	1.8
aryl-hydrocarbon receptor nuclear translocator 2	ARNT2	0.0057	2.0
calmodulin regulated spectrin-associated protein 1-like 1	CAMSAP1L1	0.0058	1.5
poly(A) binding protein, cytoplasmic 4 (inducible form)	PABPC4	0.0058	1.7
transmembrane protein with EGF-like and two follistatin-like domains 2	TMEFF2	0.0059	2.1
son of sevenless homolog 1 (Drosophila)	SOS1	0.0059	1.6
transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	TLE1	0.0060	1.5
ubiquitin-conjugating enzyme E2K	UBE2K	0.0061	1.5

(UBC1 homolog, yeast)			
clusterin	CLU	0.0061	1.6
scavenger receptor class A, member 3	SCARA3	0.0061	1.8
putative homeodomain transcription factor 2	PHTF2	0.0062	2.0
transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	TLE1	0.0062	2.2
ubiquitin associated and SH3 domain containing, B	UBASH3B	0.0063	2.6
hypothetical protein LOC221710	LOC221710	0.0063	1.7
breast cancer metastasis-suppressor 1-like	BRMS1L	0.0063	1.9
metallothionein 1E	MT1E	0.0063	6.0
interferon responsive gene 15	IFRG15	0.0064	1.7
Ras association (RalGDS/AF-6) domain family member 5	RASSF5	0.0064	1.8
sperm specific antigen 2	SSFA2	0.0065	2.8
WW and C2 domain containing 2	WWC2	0.0065	1.7
CDNA clone IMAGE:5263531	---	0.0066	1.5
ubiquitin associated and SH3 domain containing, B	UBASH3B	0.0066	2.0
chromosome 8 open reading frame 47	C8orf47	0.0066	2.6
OTU domain containing 4	OTUD4	0.0067	1.7
dermatan sulfate epimerase	DSE	0.0067	4.2
Lix1 homolog (mouse)-like	LIX1L	0.0067	2.1
brain-derived neurotrophic factor	BDNF	0.0068	3.0
myosin VA (heavy chain 12, myoxin)	MYO5A	0.0068	3.9
carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7	CHST7	0.0068	1.8
katanin p60 subunit A-like 1	KATNAL1	0.0068	1.5
pecanex homolog (Drosophila)	PCNX	0.0069	1.5
solute carrier family 25, member 37	SLC25A37	0.0069	2.0
wings apart-like homolog (Drosophila)	WAPAL	0.0069	1.7
DPH3, KTI11 homolog (S. cerevisiae)	DPH3	0.0071	1.7
paraneoplastic antigen MA1	PNMA1	0.0071	4.1
transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	TLE1	0.0072	2.2
serine/threonine kinase 17b	STK17B	0.0072	2.1
toll-like receptor 3	TLR3	0.0073	1.7
cystatin E/M	CST6	0.0073	3.9
tripartite motif-containing 33	TRIM33	0.0074	1.5
breast cancer metastasis-suppressor 1-like	BRMS1L	0.0075	1.5
NFKB inhibitor interacting Ras-like 1	NKIRAS1	0.0076	2.3
CDNA FLJ39093 fis, clone NT2RP7020112	---	0.0076	1.7
metallothionein 2A	MT2A	0.0076	4.1
solute carrier family 25, member 37	SLC25A37	0.0077	1.8
KIAA1946	KIAA1946	0.0077	1.8
protein tyrosine phosphatase, non-receptor type 14	PTPN14	0.0077	1.9
WW and C2 domain containing 2	WWC2	0.0078	2.2
sperm specific antigen 2	SSFA2	0.0080	1.8
aldehyde dehydrogenase 1 family, member A3	ALDH1A3	0.0080	4.5
high mobility group AT-hook 2	HMGA2	0.0081	1.7
homeobox A5	HOXA5	0.0082	3.0
myosin VA (heavy chain 12, myoxin)	MYO5A	0.0082	2.6

TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa	LOC728198 /// TAF9B	0.0083	1.7
suppressor of cytokine signaling 5	SOCS5	0.0085	1.6
similar to RIKEN cDNA 2310016C16	LOC493869	0.0085	5.0
hypothetical protein LOC221710	LOC221710	0.0085	1.9
Transcribed locus	---	0.0087	1.5
neurofibromin 2 (merlin)	NF2	0.0087	2.0
heparanase	HPSE	0.0087	2.7
PDZ and LIM domain 2 (mystique)	PDLIM2	0.0088	2.0
putative homeodomain transcription factor 2	PHTF2	0.0088	1.5
WD repeat domain 6	WDR6	0.0089	1.7
Transcribed locus, strongly similar to NP_068590.1 WW45 protein [Homo sapiens]	---	0.0090	2.1
cylindromatosis (turban tumor syndrome)	CYLD	0.0090	1.7
NK3 homeobox 1	NKX3-1	0.0091	2.6
dipeptidyl-peptidase 8	DPP8	0.0094	1.6
chromosome 1 open reading frame 79	C1orf79	0.0096	2.0
stearoyl-CoA desaturase 5	SCD5	0.0096	4.1
metallothionein 1 pseudogene 2	MT1P2	0.0097	2.4
nucleoporin 54kDa	NUP54	0.0098	1.7
hairy and enhancer of split 7 (Drosophila)	HES7	0.0099	1.5
epidermal retinal dehydrogenase 2	RDHE2	0.0099	3.2
suppressor of Ty 6 homolog (S. cerevisiae)	SUPT6H	0.0100	2.0
interleukin 6 signal transducer (gp130, oncostatin M receptor)	IL6ST	0.0101	1.8
leucine rich repeat containing 42	LRRC42	0.0102	1.8
RNA binding motif, single stranded interacting protein 2	RBMS2	0.0102	1.6
golgi associated, gamma adaptin ear containing, ARF binding protein 1	GGA1	0.0102	1.5
sterile alpha motif domain containing 4A	SAMD4A	0.0102	2.2
KIT ligand	KITLG	0.0102	4.7
CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2	CDS2	0.0103	1.5
R3H domain and coiled-coil containing 1	R3HCC1	0.0104	1.7
Transcribed locus, strongly similar to NP_689952.1 homeobox A9 [Homo sapiens]	---	0.0104	1.9
radixin	RDX	0.0104	2.8
polymerase I and transcript release factor	PTRF	0.0105	2.4
NMDA receptor regulated 1	NARG1	0.0105	2.1
solute carrier family 25, member 37	SLC25A37	0.0105	1.8
integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	ITGA3	0.0106	1.9
protein phosphatase 3 (formerly 2B), catalytic subunit, gamma isoform	PPP3CC	0.0106	1.5
transient receptor potential cation channel, subfamily C, member 1	TRPC1	0.0106	2.0
interleukin 6 signal transducer (gp130, oncostatin M receptor)	IL6ST	0.0106	2.0

Chromosome 6 open reading frame 204	C6orf204	0.0107	1.5
mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	0.0107	2.1
signal peptidase complex subunit 3 homolog (S. cerevisiae)	SPCS3	0.0107	1.6
basic transcription factor 3-like 4	BTF3L4	0.0107	1.5
cell division cycle 27 homolog (S. cerevisiae)	CDC27	0.0108	1.5
torsin A interacting protein 2	TOR1AIP2	0.0110	2.8
cytochrome P450, family 27, subfamily B, polypeptide 1	CYP27B1	0.0111	1.8
solute carrier family 25, member 37	SLC25A37	0.0111	1.8
sphingosine-1-phosphate lyase 1	SGPL1	0.0112	3.1
ring finger protein 11	RNF11	0.0112	1.8
ubiquitin specific peptidase 38	USP38	0.0112	1.7
metallothionein 1F	MT1F	0.0112	3.3
metallothionein 1E /// metallothionein 1H /// metallothionein 1M /// metallothio	MT1E /// MT1H /// MT1M /// MT1P2	0.0113	3.3
transforming growth factor, alpha	TGFA	0.0116	1.5
ataxin 3	ATXN3	0.0116	1.6
solute carrier family 25, member 37	SLC25A37	0.0116	2.2
PDZ and LIM domain 7 (enigma)	PDLIM7	0.0117	1.9
tetratricopeptide repeat domain 7B	TTC7B	0.0117	2.3
breast carcinoma amplified sequence 4	BCAS4	0.0117	1.6
polycystic kidney disease 2 (autosomal dominant)	PKD2	0.0117	2.1
cofilin 2 (muscle)	CFL2	0.0117	3.3
PBX/knotted 1 homeobox 1	PKNOX1	0.0118	1.5
V-raf-1 murine leukemia viral oncogene homolog 1	RAF1	0.0119	1.5
NGFI-A binding protein 2 (EGR1 binding protein 2)	NAB2	0.0119	1.6
PTPRF interacting protein, binding protein 1 (liprin beta 1)	PPFIBP1	0.0120	2.4
CDNA clone IMAGE:5268630	---	0.0121	1.8
karyopherin alpha 6 (importin alpha 7)	KPNA6	0.0122	1.8
---	---	0.0122	1.5
CDNA FLJ37485 fis, clone BRAWH2014379	---	0.0122	2.4
SUMO1/sentrin specific peptidase 5	SENP5	0.0122	1.6
homeobox A9	HOXA9	0.0123	3.6
beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group)	B3GALNT1	0.0124	3.1
carboxypeptidase, vitellogenic-like	CPVL	0.0124	2.4
Fraser syndrome 1	FRAS1	0.0124	7.5
dedicator of cytokinesis 5	DOCK5	0.0125	2.1
dynein, cytoplasmic 1, light intermediate chain 1	DYNC1LI1	0.0125	1.5
Similar to RIKEN cDNA 2310016C16	LOC493869	0.0126	2.2
aminopeptidase puromycin sensitive	NPEPPS	0.0127	1.5
nuclear receptor subfamily 2, group F, member 2	NR2F2	0.0127	2.7
tubulin, beta 6	TUBB6	0.0127	5.9
GULP, engulfment adaptor PTB	GULP1	0.0127	5.3

domain containing 1			
CUG triplet repeat, RNA binding protein 1	CUGBP1	0.0128	1.5
chromosome 6 open reading frame 62	C6orf62	0.0128	1.6
IKK interacting protein	IKIP	0.0128	2.2
interleukin 6 signal transducer (gp130, oncostatin M receptor)	IL6ST	0.0129	2.0
PDZ and LIM domain 5	PDLIM5	0.0129	1.8
chromosome 1 open reading frame 58	C1orf58	0.0130	2.0
cell growth regulator with ring finger domain 1	CGRRF1	0.0130	1.6
eukaryotic translation initiation factor 2C, 3	EIF2C3	0.0131	1.7
Transcribed locus	---	0.0131	1.6
EH-domain containing 1	EHD1	0.0131	1.6
potassium channel tetramerisation domain containing 9	KCTD9	0.0131	1.9
wings apart-like homolog (Drosophila)	WAPAL	0.0132	1.7
homeobox A9	HOXA9	0.0133	5.9
complement component (3d/Epstein Barr virus) receptor 2	CR2	0.0133	2.2
CDNA FLJ37485 fis, clone BRAWH2014379	---	0.0134	1.7
adaptor-related protein complex 4, sigma 1 subunit	AP4S1	0.0134	1.6
Na+/H+ exchanger domain containing 2	NHEDC2	0.0137	1.6
neurofibromin 2 (merlin)	NF2	0.0137	1.9
SMAD specific E3 ubiquitin protein ligase 2	SMURF2	0.0137	2.1
zinc finger protein 131	ZNF131	0.0138	1.5
tight junction protein 1 (zona occludens 1)	TJP1	0.0138	2.0
coiled-coil domain containing 75	CCDC75	0.0138	1.6
tubulin, beta 2A	TUBB2A	0.0139	3.2
DnaJ (Hsp40) homolog, subfamily B, member 4	DNAJB4	0.0140	1.9
glypican 1	GPC1	0.0140	2.4
Chromosome 9 open reading frame 3	C9orf3	0.0142	2.0
dual specificity phosphatase 3	DUSP3	0.0142	1.5
chondroitin sulfate N-acetylgalactosaminyltransferase 2	CSGALNACT2	0.0142	1.8
ezrin	EZR	0.0142	2.2
glutathione S-transferase theta 2	GSTT2	0.0142	1.9
RUN and FYVE domain containing 2	RUFY2	0.0143	1.7
ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)	ARIH1	0.0143	1.6
homeobox A7	HOXA7	0.0143	2.5
tubulin, alpha 1a	TUBA1A	0.0143	3.2
platelet-derived growth factor receptor-like	PDGFRL	0.0144	1.8
ezrin	EZR	0.0145	2.3
CDNA FLJ25731 fis, clone TST05584	---	0.0145	1.6
chromosome 14 open reading frame 102	C14orf102	0.0145	1.5
MAP7 domain containing 1	MAP7D1	0.0146	1.9
muskelin 1, intracellular mediator	MKLN1	0.0146	1.5

containing kelch motifs			
eukaryotic translation initiation factor 2C, 1	EIF2C1	0.0146	1.6
Rho-associated, coiled-coil containing protein kinase 1	ROCK1	0.0146	1.7
mitochondrial carrier triple repeat 1	MCART1	0.0147	1.5
major histocompatibility complex, class I, B /// major histocompatibility comple	HLA-B /// HLA-C /// MICA /// MICB /// XXbac-BPG181B23.1	0.0148	1.6
heparanase	HPSE	0.0148	2.0
Transcribed locus	---	0.0149	1.8
suppressor of cytokine signaling 6	SOCS6	0.0149	1.7
pannexin 1	PANX1	0.0149	1.9
nucleoredoxin	NXN	0.0150	4.7
homeobox A7	HOXA7	0.0150	1.6
brain-derived neurotrophic factor	BDNF	0.0151	1.7
UL16 binding protein 2	ULBP2	0.0152	1.6
Early growth response 1	EGR1	0.0152	2.0
---	---	0.0152	1.9
NGFI-A binding protein 2 (EGR1 binding protein 2)	NAB2	0.0152	1.7
major histocompatibility complex, class I, B /// major histocompatibility comple	HLA-B /// HLA-C /// MICA /// MICB /// XXbac-BPG181B23.1	0.0153	1.8
TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa	LOC728198 /// TAF9B	0.0153	2.3
CCR4-NOT transcription complex, subunit 7	CNOT7	0.0153	2.1
kinesin family member 1B	KIF1B	0.0153	1.9
Mdm4 p53 binding protein homolog (mouse)	MDM4	0.0153	1.6
cysteine-rich, angiogenic inducer, 61	CYR61	0.0154	3.8
CD55 molecule, decay accelerating factor for complement (Cromer blood group)	CD55	0.0155	2.8
chromosome 10 open reading frame 137	C10orf137	0.0155	1.6
PDZ and LIM domain 5	PDLIM5	0.0156	1.6
Transcribed locus	---	0.0156	1.6
leucine zipper-EF-hand containing transmembrane protein 2	LETM2	0.0156	1.9
sphingomyelin synthase 2	SGMS2	0.0157	1.8
integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes	ITGB1	0.0157	2.0
ubiquitin-like modifier activating enzyme 6	UBA6	0.0158	1.5
ubiquitin protein ligase E3A (human papilloma virus E6-associated protein, Angel	UBE3A	0.0158	1.5
ezrin	EZR	0.0159	2.0
ubiquitin associated and SH3 domain containing, B	UBASH3B	0.0159	2.1
glycerophosphodiester phosphodiesterase domain containing 3	GDPD3	0.0160	1.9
Rho-associated, coiled-coil containing protein kinase 1	ROCK1	0.0160	1.5

Transcribed locus	---	0.0160	1.5
beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group)	B3GALNT1	0.0161	3.4
karyopherin alpha 5 (importin alpha 6)	KPNA5	0.0162	1.6
torsin A interacting protein 2	TOR1AIP2	0.0163	1.7
protein phosphatase 1, catalytic subunit, beta isoform	PPP1CB	0.0163	1.6
iduronate 2-sulfatase (Hunter syndrome)	IDS	0.0164	1.5
transmembrane and tetratricopeptide repeat containing 3	TMTC3	0.0164	1.9
RAB8B, member RAS oncogene family	RAB8B	0.0164	1.6
Transcribed locus	---	0.0165	1.5
Morf4 family associated protein 1-like 1	MRFAP1L1	0.0165	1.5
v-Ha-ras Harvey rat sarcoma viral oncogene homolog	HRAS	0.0166	1.8
CD55 molecule, decay accelerating factor for complement (Cromer blood group)	CD55	0.0166	2.7
oxysterol binding protein-like 8	OSBPL8	0.0167	1.5
ring finger protein 24	RNF24	0.0167	1.5
WW domain containing transcription regulator 1	WWTR1	0.0167	4.3
oxysterol binding protein-like 8	OSBPL8	0.0167	1.6
ras homolog gene family, member B	RHOB	0.0168	2.2
ras homolog gene family, member D	RHOD	0.0168	2.0
ubiquitin family domain containing 1	UBFD1	0.0169	1.7
AKT interacting protein	AKTIP	0.0169	1.6
CDNA FLJ38498 fis, clone FELNG2000241	---	0.0170	1.8
ubiquitin specific peptidase 15	USP15	0.0171	1.6
myeloid cell leukemia sequence 1 (BCL2-related)	MCL1	0.0171	1.5
zinc finger, DHHC-type containing 2	ZDHHC2	0.0172	3.3
inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	ID2	0.0172	2.4
polo-like kinase 4 (Drosophila)	PLK4	0.0173	1.7
folate receptor 1 (adult)	FOLR1	0.0173	2.1
cofilin 2 (muscle)	CFL2	0.0173	2.7
ubiquitin specific peptidase 46	USP46	0.0174	1.5
transient receptor potential cation channel, subfamily C, member 1	TRPC1	0.0174	1.7
cell division cycle 37 homolog (S. cerevisiae)-like 1	CDC37L1	0.0174	1.5
solute carrier family 25, member 30	SLC25A30	0.0175	1.6
Arsenic transactivated protein 1	---	0.0175	1.7
menage a trois homolog 1, cyclin H assembly factor (Xenopus laevis)	MNAT1	0.0175	1.7
KIT ligand	KITLG	0.0175	3.6
Armadillo repeat containing, X-linked 4	ARMCX4	0.0175	1.5
nephroblastoma overexpressed gene	NOV	0.0175	1.6
IKK interacting protein	IKIP	0.0177	1.8
SERTA domain containing 2	SERTAD2	0.0178	1.5
tRNA-yW synthesizing protein 3	TYW3	0.0179	1.9

homolog (S. cerevisiae)			
cofilin 2 (muscle)	CFL2	0.0181	2.9
inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	ID2	0.0182	1.9
chromosome 6 open reading frame 168	C6orf168	0.0183	2.0
ring finger protein 141	RNF141	0.0183	1.7
metallothionein 1F	MT1F	0.0183	2.2
ubiquitin carboxyl-terminal hydrolase L5	UCHL5	0.0183	1.5
fukutin related protein	FKRP	0.0183	1.5
caveolin 2	CAV2	0.0184	5.7
CD55 molecule, decay accelerating factor for complement (Cromer blood group)	CD55	0.0184	2.5
BCL2-associated athanogene 4	BAG4	0.0185	2.2
CP110 protein	CP110	0.0185	1.5
cysteine-rich, angiogenic inducer, 61	CYR61	0.0186	3.2
transmembrane protein 33	TMEM33	0.0187	1.6
schwannomin interacting protein 1	SCHIP1	0.0187	2.4
ubiquilin 1	UBQLN1	0.0188	1.7
methyltransferase like 6	METTL6	0.0189	1.5
forty-two-three domain containing 1	FYTTD1	0.0191	1.5
SET domain containing (lysine methyltransferase) 7	SETD7	0.0191	1.6
CAP-GLY domain containing linker protein family, member 4	CLIP4	0.0192	1.5
Eukaryotic translation elongation factor 1 epsilon 1	EEF1E1	0.0192	1.5
forkhead box C1	FOXC1	0.0192	1.6
suppressor of cytokine signaling 6	SOCS6	0.0193	1.7
choroideremia-like (Rab escort protein 2)	CHML	0.0193	1.9
solute carrier family 4, anion exchanger, member 3	SLC4A3	0.0194	1.6
keratin associated protein 2-1 /// keratin associated protein 2-4 /// keratin as	KAP2.1B /// KRTAP2-1 /// KRTAP2-4 /// LOC644350 /// LOC728285 /// LOC728934 ///	0.0194	2.6
poly (ADP-ribose) polymerase family, member 6	PARP6	0.0195	1.6
beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group)	B3GALNT1	0.0197	3.6
chromosome 10 open reading frame 18	C10orf18	0.0197	1.7
asparagine-linked glycosylation 10 homolog (yeast, alpha-1,2-glucosyltransferase	ALG10	0.0197	1.6
junction-mediating and regulatory protein	JMY	0.0197	1.5
homeobox D10	HOXD10	0.0198	1.5
laminin, beta 2 (laminin S)	LAMB2	0.0199	1.6

Table B.2 List of genes upregulated in lumen lines, sorted by *p*-value. All genes are statistically significant at a significance level of 0.02 and are upregulated at least 1.5-fold in lumen lines.

Gene Name	Abbreviation	p-value	Fold-Change
Transcribed locus	---	0.0000	1.6
glutathione S-transferase A4	GSTA4	0.0000	2.7
chemokine (C-C motif) ligand 14 /// chemokine (C-C motif) ligand 15	CCL14 /// CCL15	0.0000	8.9
aminoacylase 1	ACY1	0.0001	2.1
Pentatricopeptide repeat domain 3	PTCD3	0.0001	1.6
apoptosis-inducing factor, mitochondrion-associated, 3	AIFM3	0.0001	4.5
malate dehydrogenase 2, NAD (mitochondrial)	MDH2	0.0002	1.8
glutathione S-transferase A4	GSTA4	0.0002	1.6
TIP41, TOR signaling pathway regulator-like (S. cerevisiae)	TIPRL	0.0002	2.1
hypothetical protein MGC3205 /// transmembrane protein 205	MGC3205 /// TMEM205	0.0002	1.7
hypothetical LOC730101	LOC730101	0.0002	2.1
quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (	QPRT	0.0003	4.3
monoacylglycerol O-acyltransferase 3	MOGAT3	0.0004	2.5
CDNA FLJ43660 fis, clone SYNOV4004823	---	0.0004	3.2
filamin binding LIM protein 1	FBLIM1	0.0004	1.6
NADPH oxidase 1	NOX1	0.0004	5.2
CDNA FLJ37336 fis, clone BRAMY2020412	---	0.0005	1.9
zinc finger protein 789	ZNF789	0.0005	1.5
solute carrier family 11 (proton-coupled divalent metal ion transporters), membe	SLC11A2	0.0005	1.5
NADPH oxidase 1	NOX1	0.0005	9.5
proline-rich acidic protein 1	PRAP1	0.0005	4.6
family with sequence similarity 105, member A	FAM105A	0.0005	4.6
mitogen-activated protein kinase kinase 6	MAP2K6	0.0005	3.6
Transcribed locus	---	0.0006	2.6
caudal type homeobox 1	CDX1	0.0006	4.3
proteasome (prosome, macropain) subunit, beta type, 7	PSMB7	0.0006	1.7
B-cell CLL/lymphoma 11B (zinc finger protein)	BCL11B	0.0006	3.4
Transcribed locus	---	0.0006	2.1
pleckstrin homology domain containing, family J member 1	PLEKHJ1	0.0007	1.5
MRNA; cDNA DKFZp686K0394 (from clone DKFZp686K0394)	---	0.0007	1.6
NADPH oxidase 1	NOX1	0.0007	4.0
claudin 3	CLDN3	0.0008	3.7
checkpoint with forkhead and ring finger domains	CHFR	0.0008	2.9

isocitrate dehydrogenase 2 (NADP+), mitochondrial	IDH2	0.0008	2.7
glutathione S-transferase kappa 1	GSTK1	0.0008	2.3
carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross re	CEACAM6	0.0009	7.0
phosphoenolpyruvate carboxykinase 2 (mitochondrial)	PCK2	0.0009	2.2
Transcribed locus	---	0.0009	2.1
catenin, beta interacting protein 1	CTNNBIP1	0.0010	1.7
Rho guanine nucleotide exchange factor (GEF) 10-like	ARHGEF10L	0.0010	1.9
claudin 3	CLDN3	0.0010	4.6
transforming, acidic coiled-coil containing protein 2	TACC2	0.0011	1.8
checkpoint with forkhead and ring finger domains	CHFR	0.0011	4.3
Transcribed locus	---	0.0011	2.1
hypothetical protein LOC257407	LOC257407	0.0011	2.1
zinc finger protein 789	ZNF789	0.0011	1.7
Transcribed locus	---	0.0012	2.1
hydroxysteroid (17-beta) dehydrogenase 4	HSD17B4	0.0013	1.8
hypothetical LOC644873	FLJ33630	0.0014	1.9
chemokine (C-C motif) ligand 14 /// chemokine (C-C motif) ligand 15	CCL14 /// CCL15	0.0014	1.6
histamine N-methyltransferase	HNMT	0.0014	3.7
carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross re	CEACAM6	0.0014	7.5
mitochondria-associated protein involved in granulocyte-macrophage colony-stimul	CORO7 /// Magmas	0.0014	1.6
CDNA clone IMAGE:5263455	---	0.0014	2.8
quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (	QPRT	0.0015	3.7
dehydrogenase/reductase (SDR family) member 1	DHRS1	0.0015	1.7
MRNA full length insert cDNA clone EUROIMAGE 375854 /// MRNA full length insert	---	0.0016	2.4
Transcribed locus	---	0.0016	1.8
Solute carrier family 25, member 46	SLC25A46	0.0016	1.5
ATPase, Na+/K+ transporting, alpha 1 polypeptide	ATP1A1	0.0017	1.6
hypothetical protein LOC257407	LOC257407	0.0017	1.8
glutathione synthetase	GSS	0.0017	1.6
MRNA full length insert cDNA clone EUROIMAGE 375854 /// MRNA full length insert	---	0.0017	2.3
hypothetical protein LOC339290	LOC339290	0.0018	2.7
Transcribed locus	---	0.0018	1.5
nth endonuclease III-like 1 (E. coli)	NTHL1	0.0018	1.5
neighbor of BRCA1 gene 2	NBR2	0.0018	2.0
solute carrier family 22, member 18	SLC22A18	0.0018	2.2
NADPH oxidase 1	NOX1	0.0019	3.8
villin 1	VIL1	0.0019	4.3

ATPase, Cu++ transporting, beta polypeptide	ATP7B	0.0019	2.6
kinesin family member 9	KIF9	0.0019	1.5
calcium and integrin binding 1 (calmyrin)	CIB1	0.0019	1.6
Transcribed locus	---	0.0020	1.8
vav 3 guanine nucleotide exchange factor	VAV3	0.0020	5.7
sialic acid acetylesterase	SIAE	0.0021	2.4
Transcribed locus	---	0.0021	1.5
filamin binding LIM protein 1	FBLIM1	0.0021	2.0
forkhead box A2	FOXA2	0.0023	3.5
syntaxin binding protein 6 (amisyn)	STXBP6	0.0023	2.9
MRNA full length insert cDNA clone EUROIMAGE 826033	---	0.0023	2.6
B-cell CLL/lymphoma 11B (zinc finger protein)	BCL11B	0.0023	2.4
dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2	DYRK2	0.0023	1.6
mucin 3B, cell surface associated	MUC3B	0.0024	2.9
mucin 3A, cell surface associated	MUC3A	0.0024	2.7
frizzled homolog 5 (Drosophila)	FZD5	0.0024	2.6
Transcribed locus	---	0.0024	1.6
erythrocyte membrane protein band 4.1 (elliptocytosis 1, RH-linked)	EPB41	0.0024	1.6
aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)	AKR7A3	0.0024	1.7
Transcribed locus	---	0.0025	1.5
MRNA from chromosome 5q21-22, clone:357Ex	---	0.0025	1.8
chromosome 18 open reading frame 18	C18orf18	0.0026	2.5
DENN/MADD domain containing 1A	DENND1A	0.0026	1.7
fatty acid binding protein 1, liver	FABP1	0.0027	17.6
cytochrome P450, family 2, subfamily S, polypeptide 1	CYP2S1	0.0027	2.8
endoplasmic reticulum aminopeptidase 1	ERAP1	0.0027	2.6
WNK lysine deficient protein kinase 4	WNK4	0.0028	3.4
isocitrate dehydrogenase 2 (NADP+), mitochondrial	IDH2	0.0028	2.6
solute carrier family 35 (UDP-glucuronic acid/UDP-N-acetylgalactosamine dual tra	SLC35D1	0.0028	1.8
Transcribed locus	---	0.0028	2.4
hephaestin	HEPH	0.0028	8.2
vascular endothelial growth factor A	VEGFA	0.0029	1.6
phosphatidylinositol glycan anchor biosynthesis, class N	PIGN	0.0029	1.9
caspase 6, apoptosis-related cysteine peptidase	CASP6	0.0029	2.0
forkhead box A2	FOXA2	0.0029	3.1
CDNA FLJ44346 fis, clone TRACH3005808, highly similar to Homo sapiens p150	---	0.0030	1.6
histamine N-methyltransferase	HNMT	0.0030	1.7
neuronal guanine nucleotide exchange factor	NGEF	0.0030	2.0

chromosome 9 open reading frame 123	C9orf123	0.0030	1.7
CDNA FLJ30340 fis, clone BRACE2007411	---	0.0031	1.6
Male sterility domain containing 1	MLSTD1	0.0031	3.1
CDNA FLJ37936 fis, clone CTONG2005468	---	0.0032	4.0
lamin B receptor	LBR	0.0033	1.5
interferon regulatory factor 8	IRF8	0.0033	3.8
vav 3 guanine nucleotide exchange factor	VAV3	0.0033	5.3
trafficking protein particle complex 6A	TRAPPC6A	0.0033	2.2
calpastatin	CAST	0.0033	2.2
zinc finger protein 775	ZNF775	0.0034	1.5
chromosome 14 open reading frame 112	C14orf112	0.0034	1.7
5-hydroxytryptamine (serotonin) receptor 1D	HTR1D	0.0034	1.5
dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	DPP4	0.0034	2.6
guanylate cyclase 2C (heat stable enterotoxin receptor)	GUCY2C	0.0035	6.5
tubulointerstitial nephritis antigen	TINAG	0.0035	2.8
mucin 3A, cell surface associated	MUC3A	0.0035	2.1
potassium voltage-gated channel, shaker-related subfamily, beta member 2	KCNAB2	0.0036	2.0
cold inducible RNA binding protein	CIRBP	0.0036	1.5
radial spoke head 1 homolog (Chlamydomonas)	RSPH1	0.0036	2.2
growth hormone regulated TBC protein 1	GRTP1	0.0037	2.6
ASCL830	UNQ830	0.0037	1.5
hypothetical protein LOC730102	LOC730102	0.0038	5.1
Transcribed locus	---	0.0038	2.1
protease, serine, 8	PRSS8	0.0039	2.7
vascular endothelial growth factor A	VEGFA	0.0039	1.9
BCL2-like 14 (apoptosis facilitator)	BCL2L14	0.0039	2.9
single-minded homolog 2 (Drosophila)	SIM2	0.0040	1.5
myosin VB	MYO5B	0.0040	1.8
hypothetical LOC400642	LOC400642	0.0041	1.7
SFT2 domain containing 3	SFT2D3	0.0041	1.5
succinate-CoA ligase, GDP-forming, beta subunit	SUCLG2	0.0041	1.7
transmembrane protein 117	TMEM117	0.0042	1.5
ARV1 homolog (S. cerevisiae)	ARV1	0.0043	2.3
X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	XPNPEP1	0.0044	1.5
Fc fragment of IgG, receptor, transporter, alpha	FCGRT	0.0044	3.5
chromosome 6 open reading frame 130	C6orf130	0.0044	1.6
chromosome 1 open reading frame 213	C1orf213	0.0045	1.5
spermatogenesis associated 2	SPATA2	0.0045	1.9
endoplasmic reticulum-golgi intermediate compartment (ERGIC) 1	ERGIC1	0.0045	1.6
anterior gradient homolog 2 (Xenopus	AGR2	0.0045	3.1

laevis)			
filamin binding LIM protein 1	FBLIM1	0.0047	2.0
abhydrolase domain containing 14B	ABHD14B	0.0047	2.1
hypothetical protein LOC129293	LOC129293	0.0048	3.5
villin 1	VIL1	0.0049	4.2
transforming, acidic coiled-coil containing protein 2	TACC2	0.0049	1.5
mature T-cell proliferation 1 /// similar to mature T-cell proliferations 1	LOC100133946 /// MTCP1	0.0050	1.6
hypothetical protein LOC728769	LOC728769	0.0050	1.5
vav 3 guanine nucleotide exchange factor	VAV3	0.0051	4.0
inositol 1,4,5-triphosphate receptor, type 2	ITPR2	0.0051	2.0
methyltransferase like 7B	METTL7B	0.0051	1.6
neuronal guanine nucleotide exchange factor	NGEF	0.0051	1.7
dopa decarboxylase (aromatic L-amino acid decarboxylase)	DDC	0.0051	5.7
hypothetical protein LOC339290	LOC339290	0.0051	1.8
caspase 6, apoptosis-related cysteine peptidase	CASP6	0.0053	1.5
hypothetical protein LOC100128822	LOC100128822	0.0053	2.8
CDNA FLJ31475 fis, clone NT2NE2001598	---	0.0054	1.5
succinate-CoA ligase, GDP-forming, beta subunit	SUCLG2	0.0055	1.7
succinate-CoA ligase, GDP-forming, beta subunit	SUCLG2	0.0055	1.7
acyl-Coenzyme A oxidase 2, branched chain	ACOX2	0.0056	3.3
DAZ associated protein 2	DAZAP2	0.0056	1.5
KIAA0152	KIAA0152	0.0056	1.8
glycogenin 2	GYG2	0.0056	2.2
proteasomal ATPase-associated factor 1	PAAF1	0.0056	1.6
ribokinase	RBKS	0.0056	1.5
ATPase, Ca++ transporting, type 2C, member 2	ATP2C2	0.0057	2.2
hydroxysteroid dehydrogenase like 2	HSDL2	0.0057	1.5
fatty acid amide hydrolase	FAAH	0.0057	1.6
hypothetical LOC400642	LOC400642	0.0058	2.0
suppressor of defective silencing 3 homolog (S. cerevisiae)	SUDS3	0.0058	1.8
lectin, galactoside-binding, soluble, 4 (galectin 4)	LGALS4	0.0059	10.0
ribosomal protein L22	RPL22	0.0060	1.5
Nebulette	NEBL	0.0060	1.6
monoamine oxidase A	MAOA	0.0061	4.7
chromosome 16 open reading frame 14	C16orf14	0.0061	1.6
tripartite motif-containing 2	TRIM2	0.0061	2.6
glutathione synthetase	GSS	0.0063	1.6
KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	KDELR2	0.0063	1.6
endoplasmic reticulum aminopeptidase 1	ERAP1	0.0063	2.0
protor-2	FLJ14213	0.0065	2.2
transmembrane protease, serine 2 ///	PP9284 /// TMPRSS2	0.0065	2.9

hypothetical protein LOC100130534			
MOB1, Mps One Binder kinase activator-like 2B (yeast)	MOBKL2B	0.0066	2.1
phosphatidylinositol glycan anchor biosynthesis, class T	PIGT	0.0066	1.7
clarin 3	CLRN3	0.0066	8.8
glycogenin 2	GYG2	0.0066	1.9
Chromosome 10 open reading frame 75	C10orf75	0.0067	1.9
FH2 domain containing 1	FHDC1	0.0068	1.5
---	---	0.0068	3.4
chromosome 20 open reading frame 118	C20orf118	0.0069	1.9
jumonji, AT rich interactive domain 1C	JARID1C	0.0069	1.5
Transcribed locus	---	0.0069	1.5
4-hydroxyphenylpyruvate dioxygenase-like	HPDL	0.0069	3.3
cold inducible RNA binding protein	CIRBP	0.0070	1.5
ATP-binding cassette, sub-family B (MDR/TAP), member 1	ABCB1	0.0071	2.6
endoplasmic reticulum-golgi intermediate compartment (ERGIC) 1	ERGIC1	0.0072	2.1
male sterility domain containing 1	MLSTD1	0.0073	2.2
RAD50 homolog (S. cerevisiae)	RAD50	0.0074	1.6
solute carrier family 39 (metal ion transporter), member 5	SLC39A5	0.0074	1.8
anterior gradient homolog 3 (Xenopus laevis)	AGR3	0.0074	6.9
UDP glucuronosyltransferase 1 family, polypeptide A6	UGT1A6	0.0074	4.3
ATPase, Ca++ transporting, type 2C, member 2	ATP2C2	0.0075	1.7
v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	MYCN	0.0075	1.6
Zinc finger protein 84	ZNF84	0.0075	1.8
Eukaryotic translation initiation factor 4E binding protein 3 /// ankyrin repeat	ANKHD1 /// ANKHD1-EIF4EBP3 /// EIF4EBP3	0.0076	1.7
IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)	IMMP2L	0.0077	1.7
histamine N-methyltransferase	HNMT	0.0077	5.2
amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2	ALS2CR2	0.0077	1.8
autism susceptibility candidate 2	AUTS2	0.0077	6.6
ethylmalonic encephalopathy 1	ETHE1	0.0077	1.9
trefoil factor 3 (intestinal)	TFF3	0.0078	5.9
ferritin, heavy polypeptide 1	FTH1	0.0078	1.9
epoxide hydrolase 2, cytoplasmic	EPHX2	0.0079	3.2
nudix (nucleoside diphosphate linked moiety X)-type motif 13	NUDT13	0.0079	1.6
tumor suppressor candidate 1	TUSC1	0.0079	4.3
replication initiator 1	REPIN1	0.0079	1.7
acetyl-Coenzyme A acyltransferase 2	ACAA2	0.0081	2.1
WNK lysine deficient protein kinase 2	WNK2	0.0081	1.8
Pentatricopeptide repeat domain 3	PTCD3	0.0081	1.5
mitogen-activated protein kinase kinase 6	MAP2K6	0.0081	2.0
chromosome 2 open reading frame 34	C2orf34	0.0082	1.5

myotubularin related protein 6	MTMR6	0.0082	1.5
chromosome 15 open reading frame 48	C15orf48	0.0083	12.2
aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)	AKR7A3	0.0083	1.7
carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	0.0083	7.8
3-hydroxyisobutyryl-Coenzyme A hydrolase	HIBCH	0.0084	1.6
RGM domain family, member B	RGMB	0.0084	3.3
TIP41, TOR signaling pathway regulator-like (S. cerevisiae)	TIPRL	0.0085	1.7
splicing factor, arginine/serine-rich 11	SFRS11	0.0086	1.5
vacuolar protein sorting 13 homolog A (S. cerevisiae)	VPS13A	0.0086	1.7
cyclin B1 interacting protein 1	CCNB1IP1	0.0088	1.5
cytochrome P450, family 2, subfamily J, polypeptide 2	CYP2J2	0.0088	2.7
protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regu	PPP1R1B	0.0089	5.7
chromosome 6 open reading frame 130	C6orf130	0.0089	1.6
chromosome 6 open reading frame 223	C6orf223	0.0091	1.7
dihydrolipoamide dehydrogenase	DLD	0.0091	1.6
family with sequence similarity 81, member A	FAM81A	0.0092	1.7
inositol 1,4,5-triphosphate receptor, type 2	ITPR2	0.0092	2.7
dolichyl-phosphate (UDP-N-acetylglucosamine) N-acetylglucosaminephosphotransfera	DPAGT1	0.0093	1.7
zinc finger protein 529	ZNF529	0.0093	3.5
Peptidylprolyl isomerase D (cyclophilin D)	PPID	0.0094	1.8
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C3 (subunit 9)	ATP5G3	0.0095	1.7
glycoprotein A33 (transmembrane)	GPA33	0.0095	6.2
lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)	LSS	0.0095	1.8
microsomal glutathione S-transferase 2	MGST2	0.0095	1.8
unc-93 homolog A (C. elegans)	UNC93A	0.0096	2.4
hydroxysteroid dehydrogenase like 2	HSDL2	0.0096	1.5
transmembrane protease, serine 2	TMPRSS2	0.0097	1.7
ATP-binding cassette, sub-family C (CFTR/MRP), member 6	ABCC6	0.0097	1.6
Transcribed locus	---	0.0097	2.4
WNK lysine deficient protein kinase 4	WNK4	0.0097	1.5
eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa	EIF2S2	0.0098	1.7
TBC1 domain family, member 16	TBC1D16	0.0098	2.0
endoplasmic reticulum aminopeptidase 1	ERAP1	0.0098	2.0
ubiquitin specific peptidase 9, X-	USP9X	0.0098	1.6

linked			
chromosome 16 open reading frame 14	C16orf14	0.0098	1.5
active BCR-related gene	ABR	0.0099	1.7
Transcribed locus	---	0.0100	1.8
ribonuclease P/MRP 25kDa subunit	RPP25	0.0101	1.6
plastin 1 (I isoform)	PLS1	0.0101	1.7
alpha-methylacyl-CoA racemase /// C1q and tumor necrosis factor related protein	AMACR /// C1QTNF3	0.0102	4.0
H3 histone, family 3A /// H3 histone, family 3B (H3.3B)	H3F3A /// H3F3B	0.0103	1.6
CDC42 effector protein (Rho GTPase binding) 5	CDC42EP5	0.0104	3.8
CDNA clone IMAGE:5313062	---	0.0105	1.6
villin 1	VIL1	0.0105	1.7
ATP-binding cassette, sub-family C (CFTR/MRP), member 6	ABCC6	0.0106	1.9
CDNA FLJ45814 fis, clone NT2RP7018126	---	0.0106	1.5
LETM1 domain containing 1	LETMD1	0.0106	1.5
acyl-CoA synthetase short-chain family member 2	ACSS2	0.0107	2.4
stearoyl-CoA desaturase (delta-9-desaturase)	SCD	0.0107	1.8
phytanoyl-CoA 2-hydroxylase	PHYH	0.0107	1.7
family with sequence similarity 40, member B	FAM40B	0.0107	1.5
zinc finger protein 313	ZNF313	0.0108	1.6
solute carrier family 39 (metal ion transporter), member 5	SLC39A5	0.0108	2.1
enoyl Coenzyme A hydratase domain containing 2	ECHDC2	0.0110	2.0
phosphofructokinase, liver	PFKL	0.0110	1.5
active BCR-related gene	ABR	0.0111	1.7
alpha-methylacyl-CoA racemase /// C1q and tumor necrosis factor related protein	AMACR /// C1QTNF3	0.0111	5.1
CDNA FLJ44316 fis, clone TRACH3000548	---	0.0113	1.5
transmembrane protein 45B	TMEM45B	0.0114	3.0
Rho GTPase activating protein 26	ARHGAP26	0.0114	2.1
dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	DPP4	0.0114	3.4
apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	APOBEC1	0.0115	2.6
myosin, heavy chain 14	MYH14	0.0115	1.8
---	---	0.0116	1.5
LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	LFNG	0.0117	2.4
hypothetical protein LOC730102	LOC730102	0.0117	2.0
transmembrane and tetratricopeptide repeat containing 4	TMTC4	0.0119	1.8
agmatine ureohydrolase (agmatinase)	AGMAT	0.0119	2.2
cell division cycle 25 homolog C (S. pombe)	CDC25C	0.0121	1.5
cytochrome P450, family 39, subfamily A, polypeptide 1	CYP39A1	0.0122	2.1

X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	XPNPEP1	0.0122	2.0
Transcribed locus, strongly similar to NP_001027392.1 syntaxin binding protein 1	---	0.0123	2.3
nicotinamide nucleotide adenylyltransferase 1	NMNAT1	0.0124	1.5
CDNA clone IMAGE:3462401	---	0.0125	1.7
FGGY carbohydrate kinase domain containing	FGGY	0.0125	1.5
acyl-CoA synthetase long-chain family member 3	ACSL3	0.0126	1.8
endoplasmic reticulum-golgi intermediate compartment (ERGIC) 1	ERGIC1	0.0126	1.7
Chromosome 7 unknown mRNA	---	0.0127	1.5
dachshund homolog 1 (Drosophila)	DACH1	0.0127	7.3
mature T-cell proliferation 1 /// similar to mature T-cell proliferations 1	LOC100133946 /// MTCP1	0.0128	1.5
vascular endothelial growth factor A	VEGFA	0.0128	1.7
DnaJ (Hsp40) homolog, subfamily A, member 2	DNAJA2	0.0130	1.5
Transcribed locus	---	0.0130	2.3
hypothetical protein LOC644246	LOC644246	0.0130	2.0
adipose differentiation-related protein	ADFP	0.0130	6.4
solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	SLC6A4	0.0131	1.9
trypsin domain containing 1	TYSND1	0.0131	1.6
zinc finger protein 253	ZNF253	0.0131	2.0
CDNA FLJ26512 fis, clone KDN07513	---	0.0132	2.0
DENN/MADD domain containing 1C	DENND1C	0.0132	1.6
solute carrier family 13 (sodium/sulfate symporters), member 4	SLC13A4	0.0134	1.6
F-box and leucine-rich repeat protein 17	FBXL17	0.0135	1.5
peptidylprolyl isomerase (cyclophilin)-like 4 /// zinc finger CCCH-type containi	PPIL4 /// ZC3H12D	0.0137	2.3
nucleobindin 1	NUCB1	0.0138	1.5
solute carrier family 5 (sodium/glucose cotransporter), member 1	SLC5A1	0.0138	2.0
CDNA FLJ26557 fis, clone LNF01992	---	0.0138	2.5
solute carrier organic anion transporter family, member 2B1	SLCO2B1	0.0138	1.9
cyclin-dependent kinase 5	CDK5	0.0139	1.5
zinc finger protein 706	ZNF706	0.0139	1.6
pleckstrin homology domain containing, family H (with MyTH4 domain) member 1	PLEKHH1	0.0140	2.0
Fc receptor, IgA, IgM, high affinity	FCAMR	0.0141	1.5
UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8	B3GNT8	0.0141	1.7
GTP binding protein 4 /// isopentenyl-diphosphate delta isomerase 2	GTPBP4 /// IDI2	0.0141	1.5
transmembrane protease, serine 2	TMPRSS2	0.0143	3.9
ATP-binding cassette, sub-family C, member 6 pseudogene 1	ABCC6P1	0.0143	1.8

alpha-methylacyl-CoA racemase /// C1q and tumor necrosis factor related protein	AMACR /// C1QTNF3	0.0143	3.6
H2A histone family, member J	H2AFJ	0.0144	2.2
myosin, heavy chain 14	MYH14	0.0145	2.0
Transcribed locus	---	0.0145	2.2
protor-2	FLJ14213	0.0147	2.0
Hypothetical LOC728730	LOC728730	0.0148	1.5
Fatty acid binding protein 1, liver	FABP1	0.0149	2.0
CDNA clone IMAGE:4620359	---	0.0149	1.5
phosphatidylinositol glycan anchor biosynthesis, class L	PIGL	0.0149	1.6
dopa decarboxylase (aromatic L-amino acid decarboxylase)	DDC	0.0149	2.7
zinc finger protein 397	ZNF397	0.0150	1.6
dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	DPP4	0.0151	2.4
cyclin D2	CCND2	0.0151	10.9
KIAA0152	KIAA0152	0.0151	1.7
solute carrier family 44, member 1	SLC44A1	0.0151	2.1
CDNA FLJ42548 fis, clone BRACE3004996	---	0.0152	1.5
programmed cell death 4 (neoplastic transformation inhibitor)	PDCD4	0.0152	2.0
chromosome 20 open reading frame 112	C20orf112	0.0152	1.5
glycerate kinase	GLYCTK	0.0154	1.6
zinc finger CCCH-type containing 13	ZC3H13	0.0155	1.6
sterile alpha motif domain containing 5	SAMD5	0.0155	4.0
vacuolar protein sorting 36 homolog (S. cerevisiae)	VPS36	0.0155	1.9
transmembrane protein 45B	TMEM45B	0.0157	2.4
aspartyl-tRNA synthetase 2, mitochondrial	DARS2	0.0157	1.5
iodotyrosine deiodinase	IYD	0.0157	1.9
Hypothetical LOC400642	LOC400642	0.0158	2.6
jumonji C domain containing histone demethylase 1 homolog D (S. cerevisiae)	JHDM1D	0.0158	2.0
cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-fa	CFTR	0.0158	3.9
Clone 25028 mRNA sequence	---	0.0159	1.6
hypothetical protein LOC157860	LOC157860	0.0160	1.8
contactin associated protein-like 2	CNTNAP2	0.0160	3.0
S-adenosylhomocysteine hydrolase	AHCY	0.0160	1.5
dolichyl-phosphate mannosyltransferase polypeptide 3	DPM3	0.0162	1.6
haloacid dehalogenase-like hydrolase domain containing 3	HDHD3	0.0162	1.7
SEC16 homolog B (S. cerevisiae)	SEC16B	0.0163	2.7
phosphofructokinase, liver	PFKL	0.0163	1.5
hypothetical protein MGC3207	MGC3207	0.0164	2.4
WD repeat domain 5B	WDR5B	0.0164	1.5
dipeptidase 1 (renal)	DPEP1	0.0164	4.8
telomerase-associated protein 1	TEP1	0.0164	1.8
hypothetical protein LOC202781	LOC202781	0.0165	1.8

v-myb myeloblastosis viral oncogene homolog (avian)	MYB	0.0165	2.6
hypothetical protein LOC153546	LOC153546	0.0165	1.8
chromosome 9 open reading frame 46	C9orf46	0.0166	1.6
annexin A4	ANXA4	0.0167	2.5
hypothetical protein FLJ20699	LL22NC03-5H6.5	0.0169	1.6
dopey family member 2	DOPEY2	0.0173	1.6
amiloride binding protein 1 (amine oxidase (copper-containing))	ABP1	0.0173	5.8
envoplakin	EVPL	0.0173	1.7
major facilitator superfamily domain containing 3	MFSD3	0.0173	1.8
methylcrotonoyl-Coenzyme A carboxylase 2 (beta)	MCCC2	0.0174	1.5
H2A histone family, member J	H2AFJ	0.0174	2.3
ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)	CLN3	0.0175	1.7
Transcribed locus	---	0.0176	1.6
zinc finger with KRAB and SCAN domains 1	ZKSCAN1	0.0177	1.9
ATPase, class V, type 10B	ATP10B	0.0179	4.3
CDNA FLJ40891 fis, clone UTERU2001110	---	0.0180	1.7
phosphatidylinositol-4-phosphate 5-kinase, type I, beta	PIP5K1B	0.0180	3.4
proprotein convertase subtilisin/kexin type 9	PCSK9	0.0180	3.0
coiled-coil domain containing 56	CCDC56	0.0181	1.5
inositol 1,4,5-triphosphate receptor, type 2	ITPR2	0.0182	1.9
myosin VIIB	MYO7B	0.0182	1.8
CDNA FLJ41394 fis, clone BRCAN2026197	---	0.0182	1.7
hydroxysteroid (11-beta) dehydrogenase 2	HSD11B2	0.0183	2.1
CDNA FLJ42315 fis, clone TRACH2019661	---	0.0184	2.0
serine/threonine kinase 38 like	STK38L	0.0184	2.0
similar to HSPC323	LOC284422	0.0184	2.1
MCF.2 cell line derived transforming sequence-like	MCF2L	0.0185	1.8
carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	0.0185	2.8
TOX high mobility group box family member 3	TOX3	0.0186	5.6
glucosamine-6-phosphate deaminase 1	GNPDA1	0.0186	1.5
hexokinase domain containing 1	HKDC1	0.0188	2.9
nuclear receptor co-repressor 1 /// chromosome 20 open reading frame 191 /// sim	C20orf191 /// LOC100133918 /// NCOR1	0.0188	1.5
DEAQ box polypeptide 1 (RNA-dependent ATPase)	DQX1	0.0188	1.7
ribonuclease P/MRP 25kDa subunit	RPP25	0.0189	2.8
contactin associated protein-like 2	CNTNAP2	0.0189	3.0
proteasome (prosome, macropain)	PSMC2	0.0190	1.6

26S subunit, ATPase, 2			
cell division cycle associated 7	CDCA7	0.0191	1.8
TOX high mobility group box family member 3	TOX3	0.0191	4.4
intraflagellar transport 172 homolog (Chlamydomonas)	IFT172	0.0191	1.6
syntaxin binding protein 6 (amisyn)	STXBP6	0.0192	1.7
lysosomal-associated membrane protein 2	LAMP2	0.0192	2.0
transmembrane BAX inhibitor motif containing 4	TMBIM4	0.0192	1.6
Unc-5 homolog C (C. elegans)-like	UNC5CL	0.0193	1.5
glycerol kinase	GK	0.0193	1.7
mucin 13, cell surface associated	MUC13	0.0194	3.4
pancreatic lipase-related protein 2	PNLIPRP2	0.0194	1.7
chromosome 10 open reading frame 58	C10orf58	0.0197	2.4
myosin, heavy chain 14	MYH14	0.0198	1.8
ATPase, Ca++ transporting, ubiquitous	ATP2A3	0.0199	1.6
Cas-Br-M (murine) ecotropic retroviral transforming sequence b	CBLB	0.0199	1.7
myosin X	MYO10	0.0199	1.7
glycogenin 2	GYG2	0.0199	1.6
ATPase family, AAA domain containing 4	ATAD4	0.0200	2.3
hydroxysteroid (17-beta) dehydrogenase 8	HSD17B8	0.0200	1.6
tight junction protein 3 (zona occludens 3)	TJP3	0.0200	1.7