Cancer stem cells in colorectal cancer: a review

Matthew J Munro,1,2 Susratha K Wickremesekera,1,3 Lifeng Peng,2 Swee T Tan,1,4 Tinte Itintean1

ABSTRACT
Colorectal cancer (CRC) is the second most common cancer in women and the third most common in men. Adenocarcinoma accounts for 90% of CRC cases. There has been accumulating evidence in support of the cancer stem cell (CSC) concept of cancer which proposes that CSCs are central in the initiation of cancer. CSCs have been the focus of study in a range of cancers, including CRC. This has led to the identification and understanding of genes involved in the induction and maintenance of pluripotency of stem cells, and markers for CSCs, including those investigated specifically in CRC. Knowledge of the expression pattern of CSCs in CRC has been increasing in recent years, revealing a heterogeneous population of cells within CRC ranging from pluripotent to differentiated cells, with overlapping and sometimes unique combinations of markers. This review summarises current literature on the understanding of CSCs in CRC, including evidence of the presence of CSC subpopulations, and the stem cell markers currently used to identify and localise these CSC subpopulations. Future research into this field may lead to improved methods for early detection of CRC, novel therapy and monitoring of treatment for CRC and other cancer types.

INTRODUCTION
Colorectal cancer (CRC) accounts for 9% of all cancers globally,1 2 making it the second most common cancer in women and the third most common in men.1 Adenocarcinoma that arises from the colonic and rectal epithelium accounts for 90% of CRC cases.3 More than 70% of CRC-related deaths are caused by metastases to the liver.4 While surgery may be potentially curative, less than 25% of cases are operable with recurrence rates of up to 70%.4 Inoperable, recurrent or metastatic CRCs are generally treated by palliative chemotherapy.4

CRC confined to the wall of the intestine (stages I and II) is potentially curable because of early detection and treatment.1 It has a 5-year survival rate of between 70% and 90%, however, most countries do not have a screening programme that allows early detection and treatment.1 In contrast, the average 5-year survival for regional-stage (nodal; stage III) and distant-stage (metastatic; stage IV) CRC is approximately 50%–70% and 10%–14%, respectively.1 2 3 4 This is mainly attributed to the tumour breaching the intestinal wall and spreading via the lymphatics to lymph nodes and systemically to distant organs via the bloodstream. The incidence of CRC increases dramatically after 50 years of age, with 90% of cases falling into this age bracket.2 6

The luminal surface of the colon consists of a single layer of columnar epithelial cells which are folded to form finger-like protrusions into the lumen7 (figure 1). The spaces between these folds are known as the crypts of Lieberkühn, the functional unit of the intestine.7 The colonic epithelium contains four distinct cell lineages: enterocytes, goblet cells, endocrine cells and Paneth cells (figure 2).

Crypt base columnar cells (CBCs) are small undifferentiated cells thought to be the true intestinal stem cells which give rise to the epithelial lineages.7 8 These stem cells possess the ability to divide asymmetrically, giving rise to identical daughter cells and also transit amplifying cells which proliferate and differentiate into enterocytes, goblet cells and endocrine cells during their upward movement through the crypt.7 9 These three epithelial cell types occupy the top half of the crypt as well as the luminal epithelial surface. Paneth cells differentiate from transit amplifying cells as they move downwards to the crypt base, where they are eventually found among the stem cell/CBC cell population.7 Paneth cells may contribute to the maintenance of the stem cell niche by producing mucosal defense barriers, modulating intestinal microflora and producing growth factors and other regulatory molecules.7 10

The epithelium of normal colon has a regular shape with evenly spaced unbranched crypts (figure 3A). The crypts of CRC display chaotic architecture characterised by branching crypts, irregular shape and large amounts of desmoplastic stroma (figure 3B). High-grade adenocarcinomas have tightly packed tumour cells, and the glands comprise less than 50% of the tumour volume (figure 3C). It would appear logical that tumourigenesis occurs in the crypts of the colonic epithelium in a bottom-up fashion, either through direct mutations of CBCs or progenitor cells, during the process of differentiation.11 Early investigations based on this concept hypothesised that adenomatous neoplasia has a ‘deep-crypt’ origin, with transformed epithelium ‘flowing’ upwards and peripherally to eventually replace the normal epithelium.12 However, more recent data points to a top-down carcinogenesis model for CRC, demonstrating that both small and large colon tumours and polyps have actively dividing cells at their periphery, in randomly distributed clusters.13 In an experiment by Cole and McLaken,13 cells were visualised 2.5 hours after staining with a radioactive nucleotide to show that active proliferation occurred at the surface epithelium of polyps. Given a usual 4–6-day migration period of cells from the crypt base to the luminal surface, it was hypothesised
that tumour growth is driven by increased proliferation in the superficial epithelium. During early tumourigenesis, cells at the crypt base appear normal when dysplastic cells are first seen at the surface epithelium. These dysplastic cells have altered adenomatous polyposis coli (APC) gene expression, as well as other gene expression patterns associated with tumours. One hypothesis was that normal-appearing cells which are already transformed migrate upwards from the bottom of the crypts and accumulate at the mucosal surface. A higher number of crypt openings than crypt bases has been observed in the epithelium of polyps and tumours. This has been explained by the process of surface epithelium invagination from an accumulation of neoplastic cells, leading to these cells being located at the base of a new adenomatous gland between two pre-existing normal glands. There may also be a contribution from genetically altered cells migrating down to the crypt base to initiate cancerous growth. These two theories would both create the false assumption that tumourigenesis is initiated at the crypt base.

This review summarises the key markers currently known to be associated with cancer stem cells (CSCs) and, specifically, CSCs in CRC. There will be a focus on the functions of these proteins and their contribution to CRC, their expression patterns, and any role they may play in current and future treatment options.

**THE CSC CONCEPT**

The CSC concept hypothesises that tumour growth is driven by CSCs, a small subpopulation of cells with embryonic stem cell (ESC) characteristics. These cells divide asymetrically to produce differentiated cancer cells as well as identical daughter cells, a characteristic shared with normal ESCs. However, CSCs have lost control of replication and differentiation, which leads to tumourigenesis. CSCs are defined by their ability to both self-renew and produce differentiated cancer cells, initiate tumour growth, and proliferate without regulation, leading to abnormal growth and slow-cycling cellular turnover which renders them resistant to therapies that target rapidly replicating cells.

The loss of replicative control leads to an increased number of ESC-like cells, further contributing to tumour growth. These CSCs and their progenies can harbour a surprising amount of variability, possibly due to mutations in DNA repair genes that may have initially contributed to carcinogenesis, which in turn produces heterogeneous tumours. The bulk of the tumour, a result of unregulated CSC activity, comprises differentiated cancer cells which are unable to initiate tumour formation or self-renew, but have inherited the capacity to proliferate in an uncontrolled manner.

Two theories have been proposed to explain how CSCs arise: (1) oncogenic mutations accumulating within normal adult cells or ESCs, leading to their uncontrolled proliferation; or (2) cellular dedifferentiation into a stem-like state, which in a cancer cell would produce CSCs.
OCT4, SOX2 and NANOG are the three genes which play a dominant role in regulating pluripotency. They are transcription factors which implement autoregulation of their own expression by binding to their own promoter regions, with expression of all three documented in a range of cancers. The Krüppel-like factor 4 (KLF4) and Sal-like protein 4 (SALL4) genes also play a role in pluripotency. Markers such as c-Myc, signal-transducer and activator of transcription 3 (STAT3), aldehyde dehydrogenase (ALDH), ATP binding cassette subfamily G member 2 (ABCG2) and CD133 have well-documented associations with cancer, normal stem cells, CSCs and chemoresistance. Some markers are of specific relevance to CRC, including CD44, leucine-rich repeat-containing G protein-coupled receptor 5 (LGR5), and epithelial cell adhesion molecule (EpCAM). These are the markers which will be reviewed in more detail.

**CRC CSC MARKERS**

**OCT4**

OCT4 is known to influence embryogenesis, stem cell maintenance, tumour growth and metastasis. Though it is an important stem cell marker, its expression has also been noted in normal colonic tissue, possibly due to its expression by CBCs. Amini et al analysed five colon tumour samples and found expression of OCT4 in all of them, primarily in the cytoplasm of the colonic epithelial cells. This expression has been suggested to drive recurrence, leading to poorer clinical outcomes, presumably by preventing apoptosis. The expression of OCT4 in CRC has been reported in cells which are undergoing epithelial-mesenchymal transition (EMT), a key step in cancer progression and metastasis, and a process which increases the stem-like phenotype of the cells.

**SOX2**

SOX2 is known to prevent differentiation of neural progenitor cells. Amini et al documented SOX2 expression in both of the stage III tumours they analysed. In a study comparing normal colon, dysplastic polyps and adenocarcinomas of 20 patients, Talebi et al demonstrated a significant correlation between SOX2 expression and CRC. SOX2 is localised to the crypt cells within the epithelium. The level of SOX2 expression in CRCs has also been correlated positively with metastasis and lymph-node infiltration. SOX2 controls the expression of OCT4, and pluripotency cannot be attained without this combination of transcription factors.

**NANOG**

NANOG also influences pluripotency via transcriptional control. Although it is vital in establishing embryonic pluripotency it has been suggested that NANOG is not necessary for maintaining undifferentiated ESCs in culture. Amini et al demonstrated the presence of NANOG in all five colon tumour samples in their study. Furthermore, Ibrahim et al detected strong nuclear NANOG staining in a subpopulation of colonic epithelial cells in primary CRC. NANOG expression is regulated by the OCT4/ SOX2 complex, and these three transcription factors function together to control the expression of pluripotency-related genes. High expression of NANOG is correlated with a poor prognosis in patients with CRC. Regulation of NANOG expression is thought to be at least partly through β-catenin Wnt signalling via the β-catenin/T-cell factor transcription factor complex.

**c-Myc**

The proto-oncogene c-Myc plays a key role as a transcription factor in proliferation, apoptosis, cell growth and differentiation in normal cells, as well as stem cell renewal. It is often upregulated in cancers and is a downstream target of the Wnt and mitogen-activated protein-kinase signalling pathways, which are frequently altered in CRC.
KLF4
KLF4 is highly expressed in undifferentiated cells where it regulates both cell cycle and pluripotency, and is necessary for maintaining self-renewal capacity.\textsuperscript{19,37} KLF4 has been shown to be overexpressed by CRC CSCs and it enables tumour sphere formation.\textsuperscript{\textsuperscript{19,37}} KLF4 enhances movement of cells in the form of migration, invasion and EMT, suggesting a possible role in metastasis.\textsuperscript{38} Knockdown of KLF4 expression inhibits these capabilities, as well as reduces chemoresistance, tumour initiation and expression of CSC markers.\textsuperscript{37}

c-Myc, KLF4, OCT4 and SOX2 are the genes used by Takahashi and Yamanaka\textsuperscript{20} to reprogramme somatic mouse fibroblasts to produce the first induced pluripotent stem cells (iPSCs) in 2007. Interestingly, NANOG was not required for reprogramming in this case.\textsuperscript{20} Noting that c-Myc expression causes death in human ESCs, a similar experiment in 2007 by Yu et al\textsuperscript{38} produced iPSCs from human primary fibroblasts using OCT4, SOX2, NANOG and LIN28. OCT4 and SOX are essential, while NANOG and LIN28 are not essential for clone formation although the efficiency of reprogramming was drastically increased by NANOG and modestly increased by LIN28.\textsuperscript{38}

SALL4
Transcription factor SALL4 controls self-renewal and pluripotency in normal embryonic and adult haematopoietic stem cells.\textsuperscript{39,40} It is associated with high tumour grade and lymph node metastasis in many types of cancer.\textsuperscript{39,40} A study by Hao et al\textsuperscript{39} involving 186 patients with a range of benign and malignant tumours shows that SALL4 co-localises with β-catenin in both the nucleus and cytoplasm of the crypt epithelium, therefore implicating it in the Wnt pathway. This interaction may be what drives lymph node metastasis and aggressiveness of CRC.\textsuperscript{39}

Khales et al\textsuperscript{41} measured the copy number of SALL4 mRNA in the peripheral blood of 72 patients with CRC and 60 healthy volunteers. When comparing SALL4 to the CRC biomarker, carcino-embryonic antigen they found that SALL4 is significantly more specific and sensitive in detecting CRC. SALL4 mRNA levels in the blood were found to be significantly higher in patients with CRC than in control subjects, but lower in patients with invasive CRC than in those with local CRC.\textsuperscript{41} Once CRC reaches the adventitia, metastasis and poor prognosis are likely. Hence high SALL4 mRNA levels in the blood seem to indicate that CRC is restricted to the muscularis propria and may lead to it being a critical and accessible biomarker for detecting CRC at a stage where surgery has a greater benefit.\textsuperscript{41}

ABCG2
A member of the ATP binding cassette (ABC) superfamily of transporter proteins, ABCG2 is involved in translocation of substrate into or out of a cell.\textsuperscript{42} ABC transporters are involved in multidrug resistance, a major problem in chemotherapy.\textsuperscript{43} ABCG2 is associated with proliferation and maintenance of both normal cells and CSCs, as well as tumour formation.\textsuperscript{42} Many CRCs express ABCG2 in the cytoplasm, and this correlates with higher-stage CRC.\textsuperscript{44} However, as many as half of CRC cases express ABCG2 on the cell membrane, and this is associated with both lymph node and distant metastases.\textsuperscript{44,43} ABCG2 expression by cells at the tumour margin and by some stromal cells has been reported.\textsuperscript{45} ABCG2 may function in the efflux of and resistance to irinotecan, an important chemotherapeutic drug for CRC.\textsuperscript{43}

STAT3
STAT3 is a transcription factor which is translocated to the nucleus when it becomes phosphorylated (pSTAT3). When it is constitutively active, STAT3 leads to proliferation, angiogenesis and invasion of cancer cells.\textsuperscript{46–48} Interleukin (IL) 6 induces STAT3-mediated proliferation by causing STAT3 to be phosphorylated.\textsuperscript{46–47} This has been observed in intestinal tumour-initiating cells.\textsuperscript{39} Furthermore, IL-6 produced by the lamina propria plays a role in preventing apoptosis of intestinal epithelial cells via STAT3 activity.\textsuperscript{49} There is also evidence that STAT3 promotes forms of inflammation which are beneficial to the tumour and prevent antitumour immune responses.\textsuperscript{46} In CRC, the presence of STAT3 has been observed in the cytoplasm of crypt epithelial cells, whereas active pSTAT3 is localised to both the cytoplasm and nucleus.\textsuperscript{48}

STAT3 represents a promising target in therapy-resistant CRCs. Expression of STAT3 has been related to chemoradiotherapy resistance in CRC, and inhibiting STAT3 causes previously resistant cells to become sensitive 5-fluorouracil.\textsuperscript{46} Furthermore, inhibition of pSTAT3 causes apoptosis and cell cycle arrest in CRC cells.\textsuperscript{51}

STAT3 involvement in CRC CSCs specifically has also been investigated. Tumour-initiating ALDH+/CD133\textsuperscript{3} cells contain more pSTAT3 than ALDH+/CD133\textsuperscript{−} cells, and inhibition of STAT3 prevents tumour initiation and CSC survival.\textsuperscript{52}

CD133
CD133 is one of the more well documented stem cell markers.\textsuperscript{33–36} It is also known as the prominin-1 glycoprotein and is thought to organise cell membrane topology.\textsuperscript{53} Initial studies to characterise CRC CSCs began by isolating CD133\textsuperscript{+} cells.\textsuperscript{55} When isolated from primary colon cancer samples, these cells are capable of forming tumours in mice, remain undifferentiated when cultured in serum-free media and become more aggressive over the span of generations.\textsuperscript{53,54} Furthermore, the majority of the tumours are comprised of CD133\textsuperscript{−} cells which are incapable of tumour initiation.\textsuperscript{55} A report by Todaro et al\textsuperscript{60} shows that CD133\textsuperscript{−} cells produce IL-4 and use it to evade apoptosis, confirmed by the observation that an α-IL-4 neutralising antibody or treatment with IL-4Rα will greatly enhance the sensitivity of CD133\textsuperscript{+} cells to chemotherapeutic drugs and increase their antitumour efficacy.

CD44
CD44 has also emerged as an important marker for CRC CSCs.\textsuperscript{61} It is a cell surface glycoprotein which functions in cell-cell interactions, adhesion of the cytoskeleton to the extracellular matrix and cell migration.\textsuperscript{57–63} Transcription of CD44 is at least in part activated by β-catenin Wnt signalling, and its overexpression is an early event in the transformation of colorectal adenoma to carcinoma.\textsuperscript{57,63} In the colonic mucosa it is used as a marker of immature differentiation.\textsuperscript{63} Knockdown of CD44 prevents tumourigenesis and clonal formation.\textsuperscript{64} Furthermore, injection of just 100 CD44\textsuperscript{+} cells is sufficient to initiate tumour formation in nude mice, and single CD44\textsuperscript{−} cells form tumour spheres with stem cell characteristics which develop into tumours when xenografted into nude mice.\textsuperscript{64} The same observation has been made for CD44+/CD24\textsuperscript{−}, CD44+/CD24\textsuperscript{−} and CD44+/CD24\textsuperscript{−} cells, as well as their ability to differentiate into enterocyte, enteroendocrine and goblet cell lineages in vitro.\textsuperscript{65}

In contrast to earlier studies, Shmelkov et al\textsuperscript{66} show that CD44+/CD24\textsuperscript{−}/CD133\textsuperscript{−} cells form the most aggressive colon tumours, removing the necessity for CD133 in tumour initiation.


Review

Furthermore, they observed that CD133+ cell populations are generally CD44+/CD24−, and CD133 is occasionally present on normal differentiated colonic epithelium. They hypothesise that the CD133+ population may give rise to a more aggressive CD133− subset which retains the ability to initiate tumours. It has also been suggested that CD133 expression is correlated with cell culture conditions and hypoxia.

ALDH

ALDH is widely considered to be a marker of both normal stem cells and CSCs. ALDH is known to be important in protection of haemopoietic stem cells and may play a similar role in CSCs, which exhibit raised levels of ALDH activity.

One study has explored the expression of ALDH1 in 24 types of normal tissues and six different primary cancer types, including CRC. At the base of normal crypts in the colonic epithelium, where progenitor/stem cells reside, there is weak staining for ALDH1 in nearly all the crypt bases, with a small number of cells stained strongly positive for ALDH1. This finding is similar to that of Huang et al who observe that ALDH1+ cells are found at low abundance at crypt bases, but increase during progression from normal to APC-mutant to adenoma. Of the 67 colon tumours investigated 94% contained ALDH1+ cells, and within these specimens 38.8% of the tumour cells expressed ALDH1. A high percentage of tumour cells expressing ALDH1 correlates with poor prognoses in a range of cancers. These ALDHHigh cells display properties of CSCs, including self-renewal, in vitro and in vivo tumour growth potential, and chemoresistance. Because normal colon has low ALDH1 expression relative to colon tumours, ALDH1 has been proposed as a promising marker for identifying CRC CSCs and potential CSC-directed therapy in CRC.

A clinical study investigating CSC markers in stage II and stage III rectal cancers shows that CD44, LGR5 and CD166 are all significantly upregulated following chemoradiation, suggesting that they are markers of therapy-resistant CRC CSC populations. Patients who exhibit low ALDH prior to chemoradiation do not recur following surgical resection, however, patients with high ALDH pretreatment have a recurrence rate of 40% within 36 months postresection. This highlights the importance of ALDH in predicting the prognosis of patients receiving post-surgery chemoradiation, and a possible correlation between ALDH and the abundance of CSCs.

Cells expressing both CD44 and epithelial-specific antigen generate heterogeneous tumours, are enriched following chemotherapy, and have elevated expression and activity of ALDH1 which has been proven to mediate cyclophosphamide chemotherapy resistance in these CSCs. ALDH has been used to further clarify the role of CD133 as a stem cell marker. A study from Ma et al shows that in hepatocellular carcinoma most ALDH+ cells are also CD133+, but there are many CD133+/ALDH− cells. Furthermore, CD133+/ALDH− cells have the highest tumour-forming potential, followed by CD133+/ALDH− and then CD133−/ALDH− populations.

LGR5

LGR5 is solely expressed by cycling CBCs at the crypt base. LGR5+ cells were proven to be CBCs by their ability to differentiate into all colonic epithelial lineages within 60 days of in vitro culture. These LGR5+ cells are found among Paneth cells at the crypt base. Paneth cells produce mucosal defense components such as lysozymes and cryptdins/defensins, as well as growth factors and other regulatory molecules essential to stem cell maintenance, namely epidermal growth factor, transforming growth factor-α, Wnt3 and Notch-ligand D114.

For this reason, Paneth cells are hypothesised to be essential in maintaining the intestinal stem cell niche, and so it is logical that proposed epithelial stem cells are located among them.

Further evidence that CRC CSCs may be LGR5+ has been shown through LGR5 knockdown causing tumour regression, and recovery of the LGR5+ cell population coinciding with tumour growth and recurrence. Additionally, Wnt signalling is downregulated when LGR5 expression is lowered, and this causes premature Paneth cell differentiation and reduced tumour production.

EpCAM

EpCAM is a marker expressed solely by epithelial cells and cancers of epithelial origin. It was discovered first in colon carcinoma, and subsequently shown in other cancers. Its native functions include homotypic Ca2+-independent cell-cell adhesion in the epithelia, cell signalling, proliferation, differentiation, metastasis and migration, tumourigenesis and upregulation of Wnt target genes c-Myc and cyclins A and E. EpCAM has a basal or basolateral cell membrane expression pattern in normal epithelium, solely basolateral in well differentiated colon adenocarcinoma, and can be found in the membranes, cytoplasm and lumen of moderately differentiated colon adenocarcinoma.

Numerous studies have isolated and characterised EpCAM+/CD44+ cells from colorectal cancers. Dalerba et al found that normal colon and colon tumours both contain two populations of cells: EpCAMHigh/CD44+ and EpCAMLow/CD44−. Only the EpCAMHigh/CD44+ population is capable of forming tumours when injected into non-obese diabetic/severe combined immunoodeficiency mice. It maintains its differentiated phenotype and the morphogenic and phenotypical heterogeneity of the parent lesion, and is enriched in tumours relative to normal colon. However, each tumour contains a unique and constant ratio of the EpCAMHigh/CD44+ cell population despite intratumourous heterogeneity. Furthermore, different tumours within the same patient have similar tissue morphology, somatic mutations, expression of tumour-suppressor and immunomodulatory proteins, and overall transcription profile. This is not an unexpected finding given that lesions within one patient are likely to share a genetic background with a specific set of somatic mutations.

EpCAM is a potential diagnostic marker in some epithelial cancers, where high expression correlates with poor survival and an advanced disease state. Furthermore, due to its overexpression in adenocarcinomas, metastases and CSCs, it has been used as a target for detecting circulating or metastasising cancer cells.

CONCLUSIONS

Significant progress has been made in the last decade in the field of CRC CSCs, with a rapid increase in the knowledge of pluripotency markers such as OCT4, SOX2, NANOG, c-Myc and KLF4. Pioneering work by both Yamanaka and Thompson has demonstrated the crucial role for these transcription factors in the transformation of mature somatic cells into an induced stem cell state. This knowledge, coupled with the demonstration of these and other stem cell-associated transcription factors in tumours, including CRC, will help to refine our understanding of the role of these genes and how they can be used to identify cancer-specific CSCs. This may potentially lead to the establishment of panels of biomarkers for both diagnosis and prognosis.
as well as patient-specific tumour surveillance as a surrogate for response to treatment.

Refining our understanding of the roles of other markers, such as SALL4 as a biomarker, ABCG2 as an enabler of multidrug resistance and STAT3 in tumour initiation and inflammation, may eventually lead to a better understanding of the biological pathways that currently allow for resistance to existing therapeutic agents. This may lead to the development of novel, safer and more effective treatment for CRC and other cancers.

These discoveries are helping to clarify the CRC CSC phenotypes and enhance our ability to isolate the subpopulations of metastasising, tumour-initiating and chemoresistant cells for more reliable and in-depth interrogation. Further understanding of the heterogeneity of CRC CSCs, the expression patterns of these subpopulations of CSCs and the mechanisms of their treatment resistance may lead to novel CSC-directed therapy, potentially in combination with conventional treatments which target cancer cells within the tumour bulk.

**Take home messages**

- The cancer stem cell concept of cancer is an emerging paradigm with accumulating supporting evidence in recent years.
- Cancer stem cells in colorectal cancer are characterised by their expression of multiple markers.
- This article reviews the current knowledge in these cancer stem cell markers and future directions of this field of research.

**REFERENCES**

Review


