Wnt target gene analysis in colorectal cancer and intestinal stem cells

Laurens G. van der Flier

Printed by: Wöhrmann Print Service

Cover: In situ hybridization for *Olfm4* reveals a stem cell-restricted expression pattern in the small intestine. Back, Proboscis Monkey (*Nasalis larvatus*), in Bako National Park, Malaysia. One of the author's favorite monkey species.

ISBN: 978-90-393-4995-3

The research described in this thesis was performed at the Hubrecht Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) affiliated with the University Medical Center in Utrecht. The studies described in this thesis were supported by grants from the BSIK Innovation programme "Stem Cells in Development and Disease" (SCDD, BSIK 03038), from the Centre for Biomedical Genetics (CBG) and from the Cancer Genomics Centre/Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO). Financial support by SCDD for the publication of this thesis is gratefully acknowledged.

Wnt target gene analysis in colorectal cancer and intestinal stem cells

Identificatie van Wnt target genen in darmkanker en darmstamcellen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 10 februari 2009 des middags te 2.30 uur

door

Laurens Gerrit van der Flier

geboren op 11 april 1977

te Leidschendam

Promotor: Prof. dr. H.C. Clevers

Contents

Chapter 1:	**Introduction: stem cells, self-renewal, and differentiation in the intestinal epithelium** Annu Rev Physiol. 2008 Sep 22 [Epub ahead of print]	7
Chapter 2:	**The intestinal Wnt/TCF signature** Gastroenterology. 2007 Feb;132(2):628-32	31
Chapter 3:	**Transcriptome profile of human colorectal adenomas** Mol Cancer Res. 2007 Dec;5(12):1263-75	39
Chapter 4:	**Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells** Mol Cell Biol. 2008 Apr;28(8):2732-44	55
Chapter 5:	**Transcription factor Achaete scute-like 2 (Ascl2) controls intestinal stem cell fate** Cell, in press	71
Chapter 6:	**Mouse models to study intestinal stem cells**	101
Chapter 7:	**Function of the *Regenerating* (*Reg*) gene family in the intestine**	111
Chapter 8:	**Summarizing discussion**	123
	Samenvatting	129
	Dankwoord	133
	Curriculum vitae	134
	List of publications	135

Chapter 1

Introduction and outline of this thesis

Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium

Reprinted from: Annu Rev Physiol. 2008 Sep 22 [Epub ahead of print]

Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium

Laurens G. van der Flier and Hans Clevers

Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences & University Medical Center Utrecht, 3584 CT, Utrecht, The Netherlands; email: l.vanderflier@niob.knaw.nl; clevers@niob.knaw.nl

Key Words

Wnt, Lgr5, Notch, colorectal cancer

Abstract

The mammalian intestine is covered by a single layer of epithelial cells that is renewed every 4–5 days. This high cell turnover makes it a very attractive and comprehensive adult organ system for the study of cell proliferation and differentiation. The intestine is composed of proliferative crypts, which contain intestinal stem cells, and villi, which contain differentiated specialized cell types. Through the recent identification of Lgr5, an intestinal stem cell marker, it is now possible to visualize stem cells and study their behavior and differentiation in a much broader context. In this review we describe the identification of intestinal stem cells. We also discuss genetic studies that have helped to elucidate those signals important for progenitor cells to differentiate into one of the specialized intestinal epithelial cell types. These studies describe a genetic hierarchy responsible for cell fate commitment in normal gut physiology. Where relevant we also mention aberrant deregulation of these molecular pathways that results in colon cancer.

INTRODUCTION

Epithelium: a continuous sheet of tightly linked cells that lines both the surfaces (e.g., skin) and the inside cavities (e.g., intestine) of the body

Stem cells: undifferentiated cells, residing in a specific location (a niche) within a tissue, that can produce one or more differentiated cell types

The primary function of the intestinal tract is the digestion and absorption of nutrients. The intestinal lumen is lined with a specialized simple epithelium, which performs the primary functions of digestion and water and nutrient absorption and forms a barrier against luminal pathogens. The gut is anatomically divided into the small intestine and the colon. The small intestine can be subdivided into the duodenum, the jejunum, and the ileum. The intestinal epithelium is the most vigorously self-renewing tissue of adult mammals (1). **Figure 1*a*** shows the organization of the intestinal epithelium into crypts and villi. The four differentiated cell types that reside within the epithelium—goblet cells, enteroendocrine cells, Paneth cells, and enterocytes—are visualized through staining with specific markers (**Figure 1*b–e***).

Proliferative cells reside in the crypts of Lieberkühn, epithelial invasions into the underlying connective tissue. The crypts harbor stem cells and their progeny, transit-amplifying cells. Transit-amplifying cells spend approximately two days in the crypt, in which they divide 4–5 times before they terminally differentiate into the specialized intestinal epithelial cell types.

In the small intestine, the surface area is dramatically enlarged through epithelial protrusions called villi. Three types of differentiated epithelial cells cover these villi: the absorptive enterocytes, mucous-secreting goblet cells, and hormone-secreting enteroendocrine cells. Three days after their terminal differentiation, the cells reach the tip of the villus, undergo spontaneous apoptosis, and are shed into the gut lumen (2). Paneth cells are unusual in that

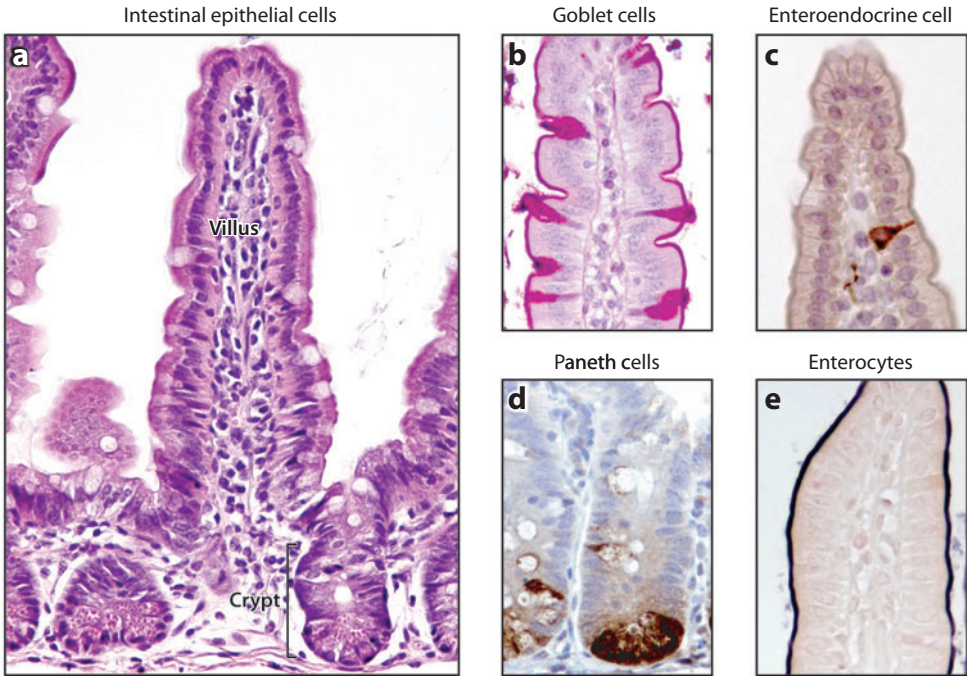


Figure 1

The intestinal epithelium. (*a*) H&E staining showing the morphology of the mouse intestine. The intestine is lined with a single layer of epithelial cells organized into invasions called the crypts of Lieberkühn and finger-like protrusions called villi. Immunohistochemical analysis for the main four differentiated cell types present in the intestinal epithelium: (*b*) periodic acid–Schiff (PAS) to stain goblet cells, (*c*) anti-synaptophysin to stain enteroendocrine cells, (*d*) lysozyme to stain Paneth cells, and (*e*) alkaline phosphatase to stain enterocytes.

they settle at the crypt bottoms and represent the only differentiated cells that escape the upward migration. Paneth cells have a function in innate immunity and antibacterial defense, to which ends they secrete bactericidal defensin peptides and lysozymes.

The modular organization of the epithelium of the small intestine and colon into crypts is globally comparable. Histologically, there are, however, two important differences between the two types of epithelia. The colon carries no villi but has a flat surface epithelium. Moreover, Paneth cells are absent in the colon.

This review focuses primarily on the physiological self-renewal of the mammalian adult small intestine. [For discussion of the development and patterning of the gut during embryogenesis, the reader is referred to excellent recent reviews (3–5).] In addition, we discuss the molecular pathways that play a role in the differentiation of stem cells into specialized epithelial cells along the crypt-villus axis in the adult intestine. Powerful genetic tools have been developed in recent years through the use of mouse transgenesis. As a consequence, the murine intestine has rapidly developed into the model of choice for the study of the intestinal epithelium. Most of the work described here is based on gain- or loss-of-function studies in genetically modified mice. The use of the Cre-LoxP system makes it possible to (in)activate genes in specific organs of interest (6). For conditional recombination in the intestinal epithelium, two Cre lines are commonly used. These are the lipophilic, xenobiotic, inducible *Cyp1a* promoter (P450) Cre line (7) and the tamoxifen-inducible villin-Cre-ER^{T2} mice (8). Upon induction, both these Cre lines are active in the intestinal epithelium, including in the intestinal stem cells.

PROLIFERATING INTESTINAL CELLS

The primary driving force behind the proliferation of epithelial cells in the intestinal crypts is the Wnt pathway. **Figure 2** shows a schematic representation of the Wnt signaling pathway.

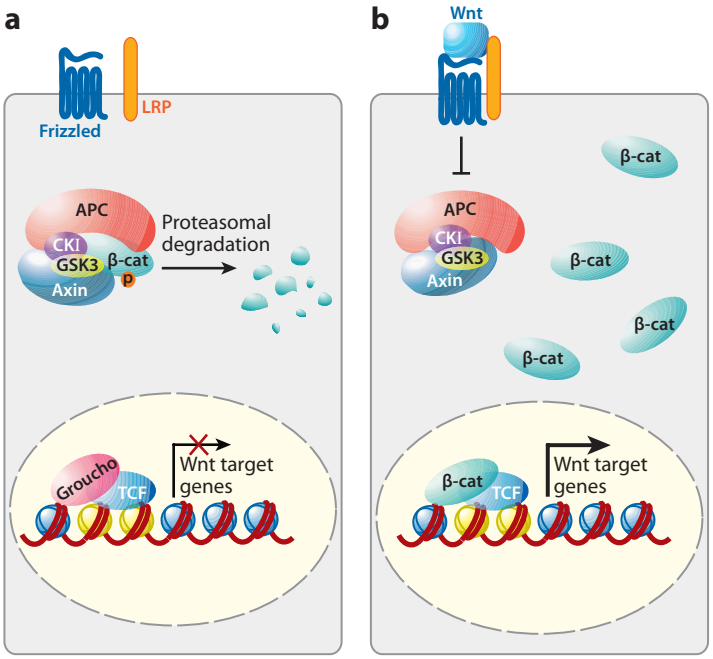


Figure 2

Schematic representation of the Wnt signaling pathway. (*a*) In the absence of Wnt stimulation, β-catenin (β-cat) levels in the cell are kept low through a dedicated destruction complex that consists of adenomatous polyposis coli (APC), casein kinase I (CKI), glycogen synthase kinase 3 (GSK3), and axin. This complex targets β-catenin for proteasomal degradation through phosphorylation. (*b*) An active Wnt signaling pathway. Wnt ligands bind to Frizzled receptors and low-density lipoprotein receptor–related protein (LRP) coreceptors. Consequently, the destruction complex no longer targets β-catenin for proteasomal degradation. The β-catenin level in the cell rises, resulting in the translocation of β-catenin into the nucleus. Here β-catenin replaces Groucho on T cell factor (TCF) transcription factors. β-Catenin/TCF form an active transcriptional complex, leading to the expression of Wnt target genes.

This pathway is highly conserved throughout the animal kingdom (9, 10). The central player in the canonical Wnt pathway is β-catenin. In the absence of a Wnt signal, β-catenin is targeted for proteasomal degradation through sequential phosphorylations occurring at its N terminus. A degradation complex, consisting of the tumor suppressors axin and adenomatous polyposis coli (APC) and the constitutively active kinases glycogen synthase kinase 3β and casein kinase I, regulates β-catenin phosphorylation status in a cell. When Wnt ligands signal through their Frizzled and low-density lipoprotein receptor–related protein (LRP) receptors, the destruction complex is inactivated

Cre-LoxP technology: a genetic method to control gene expression on the basis of short specific LoxP sequences that can site-specifically recombine through the Cre recombinase

Wnts: secreted (lipid modified) ligands for the Wnt signaling pathway; 19 different homologs exist in the mammalian genome

Canonical Wnt pathway: regulates, through a set of evolutionarily highly conserved proteins, the interaction of β-catenin with TCF/LEF transcription factors

APC: adenomatous polyposis coli

TCF/LEF: T cell factor/lymphocyte enhancer factor

CRC: colorectal cancer

in a not fully understood manner. As a result, β-catenin is no longer phosphorylated and accumulates in the cell. The coincident translocation of β-catenin into the nucleus results in the binding of β-catenin to transcription factors of the T cell factor/lymphocyte enhancer factor (TCF/LEF) family. TCF/LEF-β-catenin form an active transcriptional complex that activates target genes. In the absence of a Wnt signal, transcriptional repressors like Groucho bind TCF/LEF transcription factors (11, 12).

The Wnt pathway regulates its transcriptional target genes through TCF target sites located in promoters and/or enhancers. The optimal TCF binding site, AGATCAAAGG, is highly conserved between the four vertebrate TCF/LEF genes and *Drosophila* Tcf (13, 14). WNT/TCF reporter plasmids such as pTOPflash (15) are commonly used to measure Wnt pathway activation. These reporters consist of concatamers of the binding motif cloned upstream of a minimal promoter. A large variety of WNT/TCF target genes have been described since the discovery that this pathway represents the dominant force behind the proliferative activity of the healthy intestinal epithelium as well as behind colorectal cancer (CRC) (see below). A detailed list can be found at the Wnt home page hosted by R. Nusse (**http://www.stanford.edu/~rnusse/pathways/targets.html**).

The advent of DNA array technology has made it possible to identify Wnt target genes on a genome-wide scale in situations in which the activity of the pathway can be manipulated (16–19). These studies show extensive overlap with microarray expression studies performed on crypt-derived RNA samples in mice (20) and humans (21). These studies have shown that intestine-specific Wnt target genes are expressed in proliferative crypt progenitors as well as in CRC cells.

Recent advances in chromatin analysis utilize a cross-linking technique to capture individual transcription factors bound on their cognate DNA motifs. Subsequent immunoprecipitation of the transcription factors and their associated chromatin reveals binding events of these specific transcription factors on a genome-wide scale. The demonstration of the direct binding of TCF factors to regulatory elements in downstream genes makes it possible to distinguish whether a Wnt target is a direct or an indirect target and identifies the pertinent regulatory elements. Serial analysis of chromatin occupancy (SACO) identified 412 β-catenin-bound sites in HCT116 CRC cells (22). In another study, which used a genome-wide tiling array, analysis of TCF4-associated chromatin revealed almost 7000 TCF4-bound DNA elements in LS174T CRC cells. Testing of these TCF4-bound regions in luciferase-based reporter gene assays demonstrated that the peaks often behaved as Wnt-controlled enhancers or promoters (23). Strikingly, these two experimental studies compare well to each other but have hardly any overlap with DNA sites predicted through an algorithm-based Tcf4 binding site prediction, the enhancer element locator (EEL) bioinformatics tool (14).

Wnt target gene expression implies that the Wnt pathway is active in a gradient, with the highest activity at the crypt bottom. A recent study documents the expression pattern of all Wnts, Frizzleds, LRPs, Wnt antagonists, and TCFs in the intestinal epithelium, showing expression of multiple Wnts by the epithelial cells at the crypt bottom and implying active Wnt signaling (24).

Functional studies confirm that the Wnt pathway constitutes the master switch between proliferation and differentiation of the epithelial cells (17). Active Wnt signaling is essential for the maintenance of crypt progenitor compartments in the intestine. This is evidenced by mice lacking the Tcf4 transcription factor (25), by the conditional depletion of β-catenin from the intestinal epithelium (7, 26), and by transgenic inhibition of extracellular Wnt signaling through the secreted Dickkopf-1 Wnt inhibitor (27, 28). In all cases, a dramatic reduction of proliferative activity was observed. In the converse experiment, activating the Wnt pathway through transgenic expression of the Wnt agonist R-Spondin-1 resulted in a massive hyperproliferation of intestinal crypts (29).

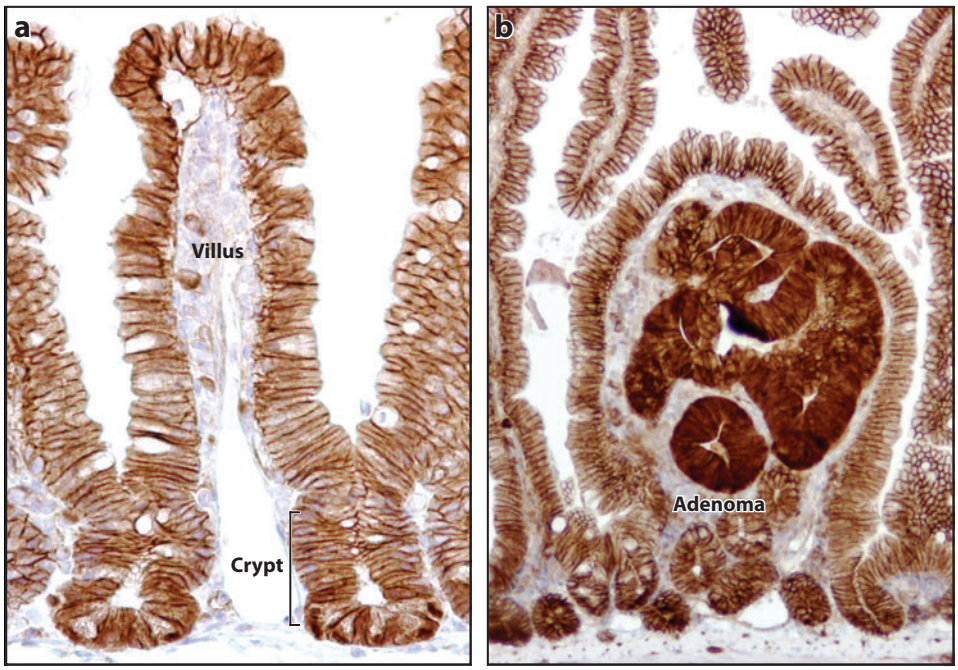


Figure 3

Active Wnt signaling in the intestine. Nuclear β-catenin is a hallmark for active Wnt signaling. (*a*) β-Catenin staining in normal wild-type intestinal epithelium. On the villus only the membranes stain positive, and at the bottom of the crypts postmitotic Paneth cells also show nuclear β-catenin staining. (*b*) Early adenoma tissue surrounded by wild-type crypt and villus epithelium in an Apc^{min} mouse shows elevated β-catenin levels in the adenoma cells, indicating hyperactive Wnt signaling.

Wnt signals in the crypt not only control proliferation of transit-amplifying progenitors but also are utilized by postmitotic Paneth cells for their terminal maturation (30). We discuss this topic in more detail below. **Figure 3** shows nuclear β-catenin staining, a hallmark for active Wnt signaling, in both the normal intestine (**Figure 3*a***) and an Apc mutant mouse adenoma (**Figure 3*b***).

Stem Cells

It has been known for decades that pluripotent stem cells fuel the proliferative activity of crypts (31, 32). Every crypt is commonly believed to contain approximately six independent stem cells; two schools of thought—the classic model and the stem cell zone model—define the exact identity of these stem cells. The small intestinal crypt has been viewed as a tube of proliferating cells limited from below by Paneth cells. Since the late 1950s, the classic model has therefore proposed that the stem cells reside at position +4 relative to the crypt bottom. Terminally differentiated Paneth cells occupy the first three positions. Potten and colleagues have championed the +4 stem cell model; they have reported that DNA label–retaining cells reside specifically at this position (33). Additionally, these researchers have observed that the +4 cells are unusually radiation-sensitive, a property proposed to functionally protect the stem cell compartment from genetic damage (34). Damaged stem cells can in this model be replaced by the first 2–3 generations of transit-amplifying cells, which would fall back into the +4 position while regaining stem cell properties. The +4 cells are actively cycling. Label retention by the +4 cell would result from asymmetric segregation of old and new DNA strands

Crypt base columnar (CBC) cells: the intestinal stem cells found at the crypt base; have Lgr5/Gpr49 expression

(34, 35). Definitive proof of stemness requires that putative stem cells be experimentally linkable to their progeny. The current literature gives no insight into the identity of the cellular progeny of +4 cells. Therefore, the position of the +4 cells in the epithelial hierarchy is not clear. The second school of thought began with the identification more than 30 years ago of crypt base columnar (CBC) cells, which are small, undifferentiated, cycling cells squeezed in between the Paneth cells (20, 36–40). Originally on the basis of morphological considerations as well as, more recently, on clonal marking techniques, Leblond, Cheng, and Bjerknes have proposed that the CBC cells may represent the crypt stem cells.

One recent method to study clonal relations within crypts involves the use of inherited methylation patterns in single adult crypts. These studies have shown that methylation increases with aging, varies between crypts, and is mosaic within single crypts. Modeling of the data predicts multiple active stem cells in human adult colon crypts (41). These stem cells are proposed to give rise to new stem cells and transit-amplifying daughters in a stochastic manner. It is predicted that approximately 95% of the time a stem cell divides asymmetrically, resulting in one stem cell and one transit-amplifying daughter. In the remaining 5% of cases, a stem cell either becomes extinct (both daughter cells differentiate) or duplicates (both daughters remain as stem cells) (42). So, in each crypt stem cells can be lost and replaced randomly in time. Entire crypts can die out. Alternatively, the intestinal epithelium can repopulate itself through crypt fission (43).

Analyses of chimeric mice confirm the notion that crypts are polyclonal but over time become monoclonal (44). The villi receive epithelial cells from multiple crypts throughout life and therefore are by definition polyclonal. A number of human studies based on natural polymorphisms or somatic mutations have provided further proof that long-lived crypts are clonal (45–47). A study by Taylor and colleagues, in contrast, has shown that mutations generating biochemical defects of cytochrome *c* oxidase create ribbons of mutant cells emanating from crypts, implying that several stem cells are simultaneously active (48). Such chimeric crypts may become monoclonal over time.

To definitively identify stem cells, one needs to be able to identify and/or mark such cells. Unfortunately, no specific marker has been available until very recently. Clonal marking of intestinal epithelial cells, through chemical mutagenesis studies, can be used to genetically mark intestinal epithelial cells by somatic mutations of the *Dlb-1* locus (49, 50). This technique was used to show that mutations in the intestinal epithelium occur in short-lived progenitors yielding one or two different cell types, as well as in long-lived cell progenitors capable of giving rise to all epithelial cells (38). Unfortunately, in this method it is not clear which cell sustains the first clonal mutation.

As described above, the Wnt pathway is the dominant force behind the proliferative activity of the intestinal epithelium both in its physiology and in CRC. For this reason, we began to study the target gene program activated by this pathway (17, 18). Through an in situ hybridization approach, we found that most of these target genes are expressed both in adenomas and in the proliferative wild-type crypts. However, a subset of these genes appeared to be restricted to a limited number of cells in the crypts (18).

One of these latter genes is *Lgr5/Gpr49*. Lgr5 is an orphan G protein–coupled receptor. Both in situ hybridization experiments as well as a LacZ knock-in allele reveal expression of Lgr5 to be restricted to the CBC cells, an observation confirmed by an enhanced green fluorescent protein (EGFP) knock-in allele (51). The EGFP in this knock-in allele is followed by CreERT2, a tamoxifen-inducible version of the Cre recombinase. A cross of this mouse strain with a strain in which the Rosa26R-LacZ reporter can be activated by Cre results in a genetic model in which tamoxifen induction genetically marks the CBC cells because the Cre induction irreversibly activates the genetic marker Rosa26R-LacZ by excision of a roadblock DNA sequence (52). **Figure 4*a***

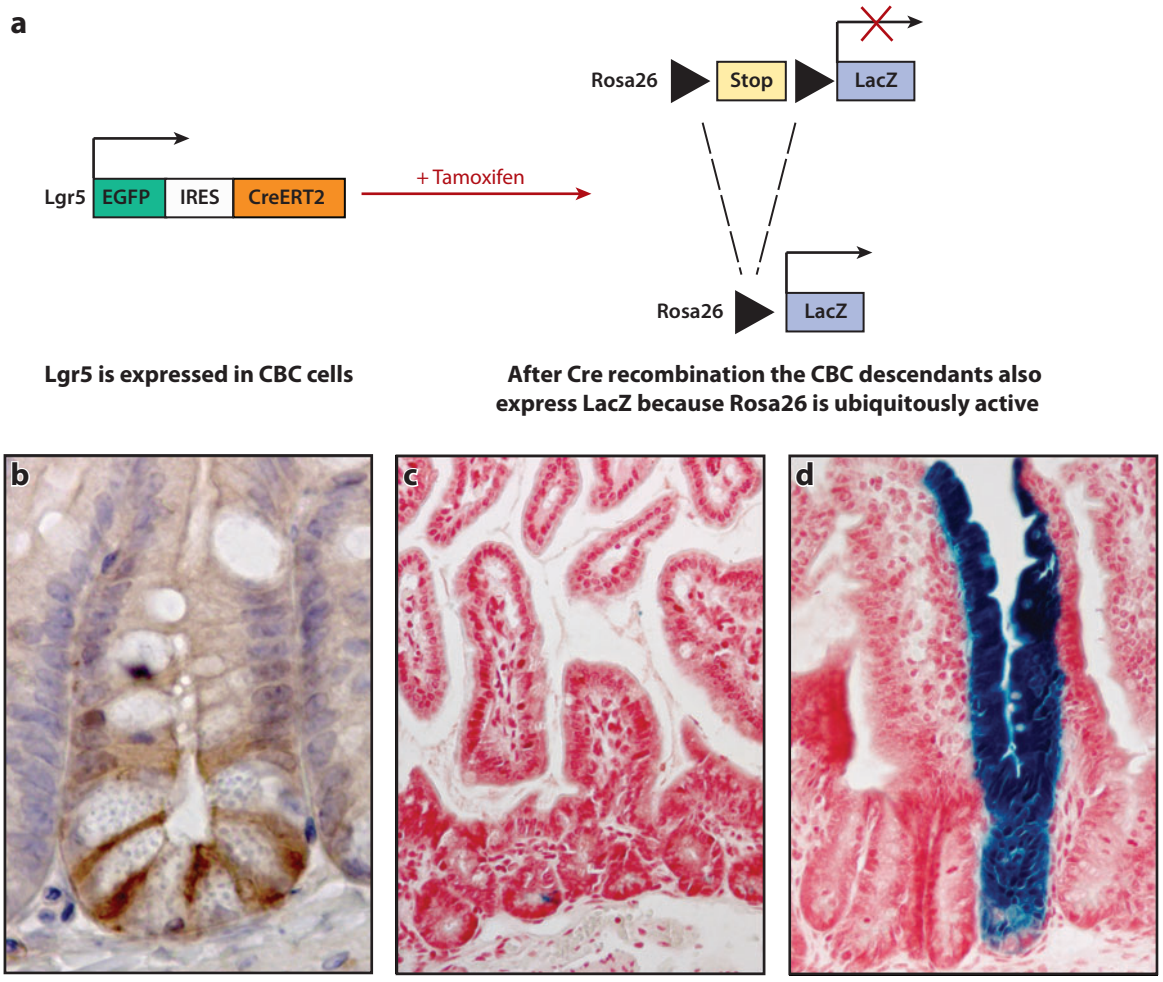


Figure 4

Lgr5 is an intestinal stem cell marker. (*a*) Genetic marking strategy to prove that Lgr5-expressing cells in the intestine are stem cells. Lgr5 knock-in mice, in which an EGFP-IRES-CreERT2 is cloned in frame with the ATG of the Lgr5 gene, are crossed with a strain in which the Rosa26R-LacZ reporter can be activated by Cre. The Rosa26R-LacZ strain by itself does not express LacZ because of a stop sequence in front of the LacZ gene. Tamoxifen-induced Cre activity in Lgr5-expressing cells will irreversibly activate LacZ by excision of the roadblock DNA sequence in the reporter strain. This genetic mark will be inherited by the daughter cells of Lgr5-positive cells because Rosa26 is ubiquitously expressed. (*b*) GFP staining of an Lgr5 knock-in allele showing that Lgr5 is expressed only in the crypt base columnar (CBC) cells that are located in between Paneth cells. (*c*) LacZ staining one day after tamoxifen induction in an Lgr5 EGFP-IRES-CreERT2 knock-in mouse crossed with a Rosa26R-LacZ reporter mice. Single cells at the bottom of the crypts, which have not had time to divide, stain for LacZ. (*d*) LacZ staining six months after a single tamoxifen induction in an Lgr5 EGFP-IRES-CreERT2 knock-in mouse crossed with a Rosa26R-LacZ reporter mice shows that crypts and adjacent villi are still lined with ribbons of LacZ-positive cells, proving that all these cells are descended from Lgr5-positive stem cells.

shows a schematic representation of this genetic lineage tracing experiment. The EGFP expression shows that Lgr5 is expressed only in CBC cells (**Figure 4*b***). Analysis at different time points after tamoxifen induction allows genetic tracing of the marked cells. Indeed, one day after Cre induction, only cells at the crypts base stain blue for LacZ (**Figure 4*c***). At later time points, blue ribbons of cells emanate from the crypts and run up the sides of the villi. These

ribbons can still be observed six months after a single tamoxifen induction (**Figure 4*d***). The clonal blue ribbons contain all four differentiated cell types of the intestine, proving that Lgr5 is a long-lived, pluripotent intestinal stem cell marker in both the small intestine and the colon (51).

Lineage tracing: the genetic marking of cells through the irreversibly excision of a roadblock DNA sequence through (inducible) Cre recombination, resulting in the expression of a marker protein (LacZ or a fluorescent protein)

Stem cell niche: the supporting microenvironment in which stem cells are found. In the intestine the niche is most likely formed by the subepithelial myofibroblasts

Intestinal Stem Cell Niche

The epithelial homeostasis of the intestine is based on a delicate balance between self-renewal and differentiation, which must be maintained throughout life. The stemness of the CBC cells in the intestine is most likely not an intrinsic property but a nonautonomous feature and is best explained by a niche model (a niche is a limited microenvironment that supports stem cells) (53). The niche for the intestinal epithelium is probably formed by the tightly associated sheath of intestinal pericryptal fibroblasts, also called subepithelial myofibroblasts (54). These cells are believed to secrete various putative growth factors and cytokines that promote epithelial proliferation (55).

Bone morphogenetic protein (BMP) signals appear to be excluded from the crypt. BMP-2 and -4 ligands are expressed in the mesenchyme of villi (56, 57). BMPs, when bound to their receptors, transduce a signal from cytoplasm to nucleus through receptor-mediated phosphorylation of SMAD transcription factors (58). The BMP signaling pathway functions as a negative regulator of crypts. Conditional deletion of Bmp receptor 1A results in hyperproliferative crypts (59). Moreover, inhibition of BMP signaling on the villus through overexpression of the BMP inhibitor Noggin results in ectopic crypt formation (56). De novo crypt formation throughout the epithelium is also seen in juvenile polyposis patients, of which at least 50% carry germ-line mutations in one of the various components of the BMP signaling pathway (60–62).

Several mice that are mutant for genes expressed in the mesenchyme show an increase in proliferative cells in their intestinal epithelium, indicating that the proteins encoded by these genes work as negative regulators of crypt cell proliferation. Examples of these mesenchymal proteins are the forkhead homolog 6 (Fkh6) (63); the homeodomain transcription factor Nkx2-3 (64); and Epimorphin, a member of the syntaxin family of membrane-bound intracellular vesicle-docking proteins (65). These phenotypic effects are possibly mediated through the BMP signaling pathway because mRNA levels of both Bmp-2 and Bmp-4 are decreased in all three null mice.

Epithelium-to-mesenchyme interactions also take place in the opposite direction. Both the hedgehog pathway and the platelet-derived growth factor-A (Pdgf-A) and its receptor (Pdgfr-α) seem to play an important role hereby. Epithelial cells express the Sonic hedgehog (Shh) and Indian hedgehog (Ihh) ligands. These ligands signal to Patched 1 and Patched 2 receptors, which are expressed by the underlying mesenchyme (66). Inhibition of the hedgehog pathway through overexpression of a pan-hedgehog inhibitor results in the reduction of villi and a hyperproliferative epithelium (66). Shh and Ihh knockouts die shortly after birth, and for this reason only intestines of animals from animals one day prior to birth have been analyzed. Both mutant mice show complex effects on embryonic gut development, with a reduced smooth muscle layer surrounding the gut and several other intestinal malformations (67). Pdgf-A, like the hedgehog ligands, is expressed by epithelial cells, although its receptor is expressed in the mesenchyme. Mice lacking Pdgf-A or Pdgfr-α develop fewer and misshapen villi (68).

CELL LINEAGE SPECIFICATION

The rare stem cells in the crypts give rise to a much larger pool of transit-amplifying cells. The transcription factor c-Myc is expressed by all cycling cells in the intestine and was the first identified target gene of the Wnt pathway in CRC cells (69). Conditional deletion of c-Myc in the intestinal epithelial results in a loss of c-Myc-deficient crypts (70). These crypts are rapidly replaced by c-Myc-proficient crypts

through crypt fission of the latter within weeks. $Myc^{-/-}$ crypt cells remain in cycle, but they are smaller, cycle slower, and divide at a smaller size compared with wild-type crypts. Myc thus appears to be essential for the biosynthetic capacity of crypt progenitor cells to successfully progress through the cell cycle (70). Strikingly, although loss of Apc in the intestinal epithelium leads to the immediate transformation of the epithelium, this phenotype is fully rescued if c-Myc is simultaneously deleted. This implies that c-Myc plays a central role in the Wnt-driven target gene program in intestinal cancer (71).

Secretory Lineages versus Absorptive Lineage: The Notch Pathway

As discussed above, transit-amplifying cells terminally differentiate into one of the four principal epithelial cell lineages of the gastrointestinal tract. Three of these cell types—the goblet cells, the enteroendocrine cells, and the Paneth cells—belong to the secretory lineage. Absorptive enterocytes represent the fourth cell type. Besides these main four cell types, researchers have described some lesser-known cell types such as, e.g., M-cells and Brush cells (72, 73). In the following paragraphs, we focus on the molecular signals that regulate the fates of the four main cell types of the intestinal epithelium. Once determined, some lineages use reiteration of signaling pathways during terminal differentiation. Other lineages appear to utilize parallel signals that are important for full mature differentiation. **Figure 5** gives a schematic overview of the genetic hierarchy in cell lineage specification in the intestine.

The Notch pathway plays a central function in these intestinal cell fate decisions. Notch genes encode single transmembrane receptors that regulate a broad spectrum of cell fate decisions and differentiation processes during animal development (74). Interaction of one of the four Notch receptors with any one of five Notch ligands results in proteolytic cleavage of the receptor within the plane of the cell membrane. A key step in the cleavage process involves the activity of the gamma-secretase protease complex. The resulting free Notch intracellular domain (NICD) translocates into the nucleus. Here NICD binds to the transcription factor RBP-Jκ (CSL or CBF1) to activate target gene transcription (75).

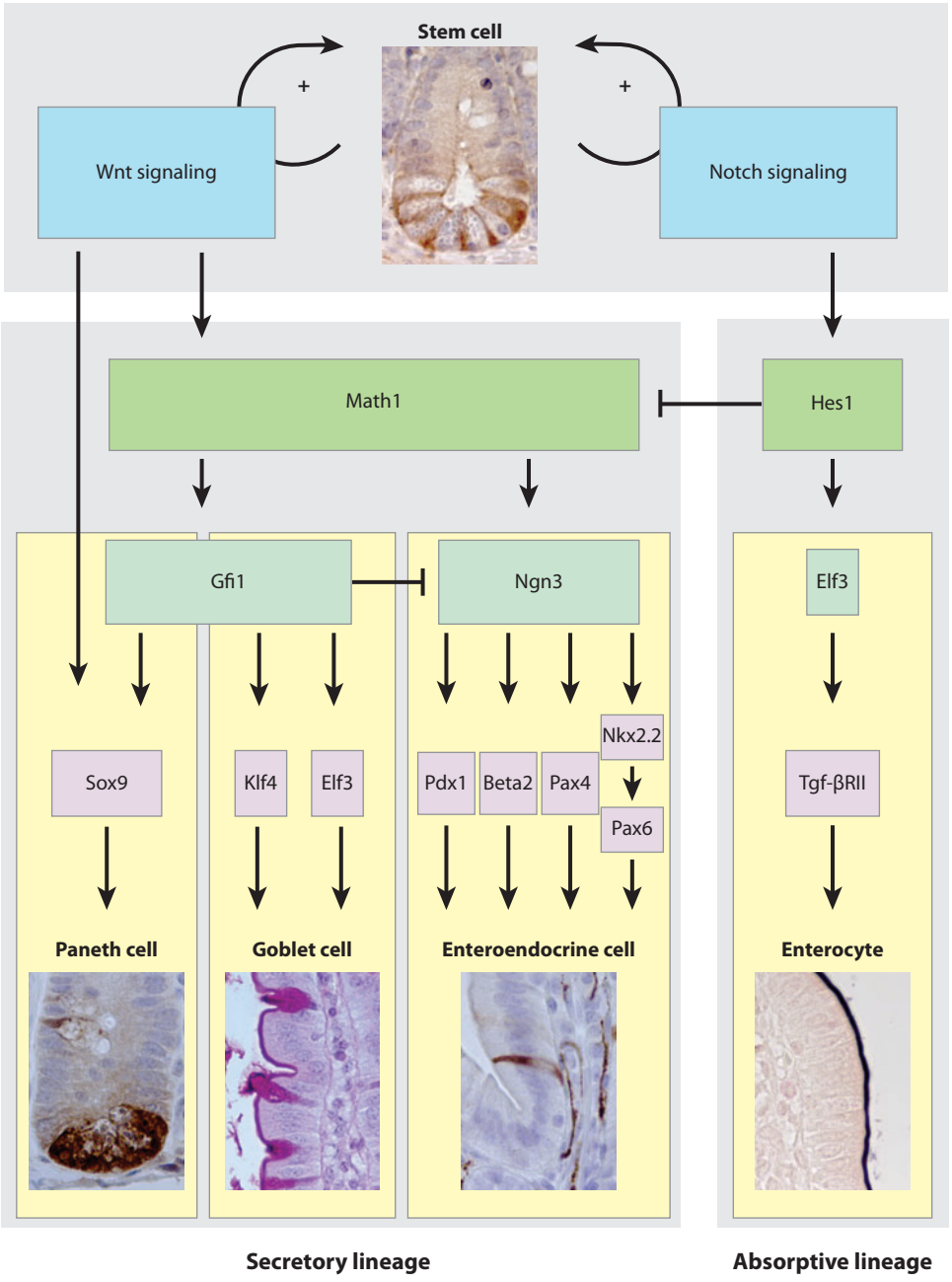
Notch pathway: a highly conserved signaling pathway involved in cell-cell communication

Like Wnt signaling, the Notch pathway is essential to maintain the crypt compartment in its undifferentiated, proliferative state. Inhibition of the Notch pathway in the intestinal epithelium by conditional deletion of the CSL gene or through pharmacological gamma-secretase inhibitors results in the rapid and complete conversion of all epithelial cells into goblet cells (76, 77). The same effect is seen in intestinal tumors in Apc^{min} mice (78). Moreover, Cre-mediated lineage tracing of Notch1 activity in the intestine shows clonal ribbons of cells, indicating that Notch1 signaling is active in adult intestinal stem cells (79). Both the Notch1 and Notch2 receptors mediate Notch signals in the intestinal epithelium. These receptors work redundantly because only conditional inactivation of both these receptors in the gut results in the complete conversion of the proliferative crypt cells into postmitotic goblet cells (80). Gain of function through specific overexpression of a constitutively active Notch1 receptor in the intestinal epithelium results in the opposite effect, a depletion of goblet cells and a reduction in enteroendocrine and Paneth cell differentiation (81, 82). Thus, the Notch pathway controls absorptive versus secretory fate decisions in the intestinal epithelium. Of note, no mutational alterations in Notch signaling in intestinal tumorigenesis have been described.

A conserved feature of Notch signaling involves the regulation of the downstream effectors of Notch. Typically, active Notch signaling results in transcription of a first tier of genes of the Hairy/Enhancer of Split (Hes) class that encode trancriptional repressors. Hes repressors in turn repress transcription of a second tier of genes, typically basic helix-loop-helix (bHLH) transcription factors that, when derepressed, induce differentiation along specific lineages. In the intestine, the direct Notch target gene

bHLH: basic helix-loop-helix

Hes1 represses transcription of the bHLH transcription factor Math1 (83). Hes1$^{-/-}$ animals are embryonic lethal, but intestines from these animals show an increase in Paneth, goblet, and enteroendocrine cells and a decrease in absorptive enterocytes (83, 84). Intestinal Math1 expression is required for commitment toward the secretory lineage because the epithelium of Math1 mutant mice is populated only by enterocytes (85). The zinc-finger transcriptional repressor Gfi1 is expressed in the secretory lineage of the intestine. Gfi1 is absent in



Math1$^{-/-}$ embryonic intestines, implying that it acts downstream of Math1. Gfi1$^{-/-}$ mice have no Paneth cells and display a clear reduction in the number of goblet cells. There is, however, an increase in the number of enteroendocrine cells (86). Mtgr1, a transcriptional corepressor, is required for secretory lineage maintenance. Mtgr1 null mice display a progressive depletion of the secretory lineage. Where Mtgr1 acts in the genetic hierarchy is not clear (for this reason Mtgr1 is not represented in **Figure 5**). The Mtgr1 null phenotype includes a loss of enteroendocrine cells, suggesting that Mtgr1 works upstream of Gfi1. However, Gfi1-positive progenitors are maintained in the absence of Mtgr1 (87).

The Wnt pathway also influences cell lineage specification of the intestine. Mouse models with impaired Wnt signaling show reductions of Math1-positive precursors in intestinal crypts. As a consequence, the secretory lineages are depleted, and the villi are lined mainly with enterocytes (7, 25, 28).

Goblet Cells

Goblet cells secrete protective mucins and trefoil proteins that are required for the movement and effective expulsion of gut contents, and provide protection against shear stress and chemical damage. The proportion of goblet cells among all epithelial cell types increases from the duodenum (~4%) to the descending colon (~16%) (88).

As discussed above, loss of function of Notch in the intestinal epithelium results in a massive conversion of epithelial cells into goblet cells (78, 89). Targeted deletion of kruppel-like factor 4 (Klf4), a zinc-finger transcription factor, results in the loss of goblet cells (90). Targeted inactivation of the most abundant secreted gastrointestinal mucin, Muc2, results in the absence of Alcian blue staining, a widely used goblet cell marker. Yet, the cells still express intestinal trefoil factor, suggesting that at least some aspects of the differentiation program of goblet cell lineage persist in Muc2$^{-/-}$ mice. Over time, Muc2 null animals frequently develop adenomas in the small intestine that progress to invasive adenocarcinomas (91).

Enteroendocrine Cells

Enteroendocrine, or neuroendocrine, cells coordinate gut functioning through specific peptide hormone secretion. There are up to 15 different subtypes of enteroendocrine cells defined by their morphology and expression of specific intestinal hormones or marker gene expression. Enteroendocrine cells are scattered as individual cells throughout the mucosa, representing approximately 1% of the cells lining the intestinal lumen (92).

As discussed above, the Notch pathway is important for the development of the secretory cell lineages of the intestine. Downstream of the Notch-Hes1-Math1 signaling cascade, Neurogenin3 (Ngn3) is required for endocrine cell fate specification. Mice homozygous for a null mutation in this bHLH transcription factor do not develop any intestinal endocrine cells (93). It is thought that, downstream of

Figure 5

Schematic overview of the genetic hierarchy of epithelial cell lineage commitment in the intestine. Intestinal stem cells proliferate under control of both the Wnt and the Notch pathway and can differentiate into all four differentiated cell types present in the intestinal epithelium. Math1 is required for the commitment to the secretory lineage. Gfi1 and Sox9 are responsible for differentiation into Paneth cells. Gfi1, kruppel-like factor 4 (Klf4), and E47-like factor 3 (Elf3) are necessary for goblet cell development. Neurogenin3 (Ngn3) is required for endocrine cell fate specification. Downstream of Ngn3, a set of transcription factors is responsible for the specification of the various enteroendocrine hormone–expressing cell types. Hairy/Enhancer of Split 1 (Hes1), through Elf3 and the transforming growth factor β type II receptor (Tgf-βRII), is responsible for differentiation into enterocytes of the absorptive lineage. The involvement of the various (transcription) factors is based on genetic experiments in mice and is discussed in more detail in the text.

Ngn3, a set of transcription factors is responsible for the specification of the various enteroendocrine hormone–expressing cell types. For secretin and cholecystokinin enteroendocrine cells, the bHLH transcription factor Beta2 (also called NeuroD) appears crucial; homozygous Beta2$^{-/-}$ mice do not have enteroendocrine cells expressing these particular hormones (94). Other genes that have been implicated in endocrine cell fate specification are the homeodomain transcription factors Pdx-1, Nkx2.2, Pax4, and Pax6 (95–98). Nkx2.2 appears to function upstream of Pax6 in enteroendocrine cell specification (95).

Paneth Cells

Paneth cells reside at the crypt base and have a function in innate immunity. They contain large apical secretory granules that contain specific proteins, including lysozymes, antimicrobials, and defensins, which likely relates to the abundance of the gastrointestinal flora. Paneth cells have a life expectancy of at least three weeks (36), much longer than that of their terminally differentiated villus counterparts. Unexpectedly, Paneth cells undergo active Wnt signaling, but they interpret these signals for their maturation and not for their proliferation (30). Labeling studies using ^{3}H-thymidine show that Paneth cells are formed around positions 5–7 of the crypt and subsequently migrate down to position 1. Thus, the Paneth cells at the bottom position have much bigger granules and are the cells showing the most degeneration (36). Paneth cells are the only differentiated intestinal epithelial cell type that migrates downward to the crypt bottom. This compartmentalization is achieved through the tyrosine kinase guidance receptors EphB2 and -3, both TCF targets. These tyrosine receptors generate repulsive forces when interacting with their ephrin-B ligands, which are expressed to high levels by villus cells. Paneth cells in both EphB3$^{-/-}$ or conditional ephrin-B1 knockout mice do not home to crypt bottoms but scatter along the crypt-villus axis (99, 100). The Frizzled-5 receptor is responsible for the Wnt activation in the Paneth cells because conditional deletion of this receptor results in EphB3-negative Paneth cells, scattered along the villus (30).

Sox9, a Wnt target (101), plays a role in the differentiation of Paneth cells in the intestinal epithelium. Conditional deletion of Sox9 in the intestinal epithelial cells results in a complete absence of Paneth cells. The Paneth cells are replaced by cycling, KI67-positive cells. Moreover, the crypts in Sox9-deficient animals are wider than in their wild-type counterparts (102, 103).

Absorptive Lineage: Enterocytes

Enterocytes, alternatively termed columnar cells, are highly polarized cells carrying an apical brush border that is responsible for absorbing and transporting nutrients across the epithelium. Enterocytes make up more than 80% of all intestinal epithelial cells.

As already mentioned above, Hes1$^{-/-}$ animals show a decrease in absorptive enterocytes (83). E47-like factor 3 (Elf3), a member of the Ets transcription family, appears to play a role in enterocyte and goblet cell differentiation. Mice homozygous for an Elf3 null mutation die shortly after birth and display poorly polarized enterocytes that have not reached maturity (104). The block in enterocyte differentiation in Elf3$^{-/-}$ mice can be genetically rescued through overexpression of transforming growth factor β type II receptor (Tgf-βRII) (105). Mice deficient for protein tyrosine kinase 6 (Ptk6) show a delay in expression of intestinal fatty acid binding protein, suggesting a role for Ptk6 in enterocyte differentiation (106).

COLORECTAL CANCER

Although Wnt signaling is very important in the normal physiology of the intestine, it was first characterized for its association with CRC, one of the most common cancers in industrialized countries. Activating mutations in the Wnt pathway initiate the overwhelming majority of CRC cases (107). These mutations either

remove the tumor suppressors APC (108–110) or axin2 (111) or activate the proto-oncogene β-catenin (112, 113). As a common result, β-catenin accumulates in the nucleus and constitutively binds to the TCF/LEF transcription factors. Tcf4 is physiologically expressed in all epithelial cells of the intestine (114). The inappropriate formation of β-catenin/TCF4 complexes results in transcriptional activation of WNT/TCF4 target genes, initiating the transformation of intestinal epithelial cells (15, 112).

Fearon & Vogelstein (115) have proposed that CRC arises through an ordered sequence of mutations in what is called the adenoma-carcinoma sequence. Invariably, the initiating mutation occurs in the Wnt pathway, leading to the formation of benign yet long-lived adenomas. Subsequently, other mutations follow, e.g., in the Kras, Smad4, and p53 genes, ultimately resulting in metastasizing carcinomas (115).

Mouse Models

Activating Wnt pathway mutations result in adenoma formation in mice. A mutagenesis study discovered the first mouse strain with hereditary adenoma formation (116). This strain was called multiple intestinal neoplasia (Min) mice and turned out to harbor a heterozygous point mutation in the Apc gene. It is thus the murine counterpart of hereditary familial adenomatous polyposis in humans. Because of spontaneous mutation of the remaining wild-type Apc allele in intestinal epithelial cells, these Min mice develop multiple adenomas in the intestine within months (117). Several Apc knockout mouse models have been generated since, with different patterns of (extra)intestinal tumor formation (118). Conditional deletion of Apc in murine intestines leads to the rapid accumulation of nuclear β-catenin and the consequent transformation of the entire intestinal epithelium (16, 119, 120).

Another way of mutationally activating the Wnt pathway involves the stabilization of β-catenin. Conditional deletion of the third exon of β-catenin, which encodes the phosphorylated residues that control its stability, also results in polyp formation in mice intestine (121). At the molecular level, the various murine models are very good models for human adenoma formation, with one distinction: The adenomas in mice are preferably formed in the small intestine, whereas in humans they are almost exclusively found in the colon both in sporadic CRC as well as in hereditary familial adenomatous polyposis. The exact reason for this difference is unknown.

Cancer Stem Cells

CRC develops from a single epithelial cell, typically mutated in its APC gene, that gains the ability to proliferate aberrantly and eventually turn malignant. One central question in colon cancer biology still to be answered is the identity of the cells that sustain the initiating mutation. Can the initiating mutation occur in any cell, or is it mandatory that a stem cell be hit? And as soon as cells are transformed, is a stem/progenitor hierarchical organization established within the newly formed tumors? In a stochastic model, every tumor cell has equal tumor-initiating potential. However, according to the cancer stem cell theory, tumors are generated and maintained by a small defined subset of undifferentiated cells able to self-renew and differentiate into the bulk tumor population. These cells are termed tumor-initiating cells or cancer stem cells (122–124). Two recent publications favor the last model. Sorting human tumor cells on the basis of CD133 expression shows that the positive fraction has up to 200-fold more tumor-initiating potential in immunodeficient mice compared with CD133-negative fractions (125, 126). The interpretation of these data may be complicated by the fact that the assay involves xenotransplantation of human tumor cells into immunodeficient mice. A recent study in which lymphomas and leukemias of mouse origin were transplanted into histocompatible mice demonstrates that most tumor cells can seed tumor growth when not challenged by a species barrier. The low frequency of tumor-sustaining cells observed in xenotransplantation studies may reflect the

limited ability of human tumor cells to adapt to growth in a foreign mouse milieu (127).

CONCLUDING REMARKS

In this review, we provide an overview and update of our current understanding of the biology of the intestinal epithelium. Recent genetic studies have created a wealth of new insights into the biology of the most rapidly self-renewing tissue of the adult mammalian body. Wnt and Notch, previously believed to uniquely control fate decisions during embryonic development, are the principal controllers of the proliferation/differentiation rheostat. Activating mutations in the Wnt cascade are the driving force behind colon cancer. A sophisticated genetic toolbox has been assembled to study in great detail the genes that control all aspects of the biology of the intestinal epithelium. These tools include a number of mouse strains that now allow the visualization, isolation, and genetic modification of the intestinal stem cells. Theories about the role of stem cells in cancer initiation and progression can now be tested.

SUMMARY POINTS

1. The intestine is lined with a rapidly self-renewing epithelium. It is organized into crypts (invaginations containing stem cells and transit-amplifying cells) and villi (protrusions that are covered with terminally differentiated cells).
2. The Wnt signaling pathway is the dominant force underlying the proliferative activity of the intestinal epithelium in both normal physiology and CRC development. The Notch pathway is needed to maintain proliferative progenitors and specific differentiated cell fates.
3. The proliferative crypt compartment of the intestine harbors CBC cells, pluripotent and long-lived stem cells that give rise to all differentiated cell types of the intestinal epithelium. The intestinal epithelium self-renews every 3–5 days.
4. Lgr5, an orphan G protein–coupled receptor, is the first definitive adult intestinal stem cell marker. Lineage tracing experiments using an inducible Cre knocked into the Lgr5 locus generate ribbons of epithelial cells of all types and genetically marked by LacZ expression for up to a year after a single Cre induction.
5. From stem cells in the crypts, a pool of transit-amplifying cells emanates. These cells will (after 2–3 days) terminally differentiate into one of the four principal epithelial cell lineages of the intestine.
6. The cancer stem cell concept proposes that tumors are generated and maintained by a small, defined subset of undifferentiated cells that are able to self-renew and differentiate into the bulk tumor population. These may be closely related to the CBC cells.

FUTURE ISSUES

1. With the identification of Lgr5 as the first definitive intestinal stem cell marker, it is now possible to visualize and isolate intestinal stem cells, study their behavior, and establish their stem cell–specific transcriptome. Novel stem cell markers should emerge from these studies.

2. The field direly lacks assays with which to study stem cells in culture or by transplantation. Such assays will be established in the near future.
3. Future efforts should be directed toward the definition of the crypt niche. It appears mandatory to identify unique marker genes for the niche cells to study their role in crypt biology and in cancer.
4. Through use of the inducible Cre recombinase integrated into the Lgr5 locus, it will be possible to genetically modify intestinal stem cells specifically for future studies. The same Cre knock-in can be used to mark Lgr5-positive cells within tumors and to study their potential cancer stem cell potential in situ.
5. An ever-increasing number of potential regulators of proliferation versus differentiation emerges from microarray studies in various models of intestinal biology. It is to be expected that—one by one—these regulators will be evaluated by what has become the standard approach, i.e., inducible deletion in the adult intestine.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Jeroen Kuipers, Nick Barker, and Hugo Snippert for their help with figure preparation.

LITERATURE CITED

1. Heath JP. 1996. Epithelial cell migration in the intestine. *Cell Biol. Int.* 20:139–46
2. Hall PA, Coates PJ, Ansari B, Hopwood D. 1994. Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. *J. Cell Sci.* 107(Pt. 12):3569–77
3. Roberts DJ. 2000. Molecular mechanisms of development of the gastrointestinal tract. *Dev. Dyn.* 219:109–20
4. Beck F, Tata F, Chawengsaksophak K. 2000. Homeobox genes and gut development. *Bioessays* 22:431–41
5. Gregorieff A, Clevers H. 2005. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* 19:877–90
6. Sauer B. 1998. Inducible gene targeting in mice using the Cre/lox system. *Methods* 14:381–92
7. Ireland H, Kemp R, Houghton C, Howard L, Clarke AR, et al. 2004. Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of β-catenin. *Gastroenterology* 126:1236–46
8. el Marjou F, Janssen KP, Chang BH, Li M, Hindie V, et al. 2004. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* 39:186–93
9. Clevers H. 2006. Wnt/β-catenin signaling in development and disease. *Cell* 127:469–80
10. Logan CY, Nusse R. 2004. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* 20:781–810
11. Cavallo RA, Cox RT, Moline MM, Roose J, Polevoy GA, et al. 1998. *Drosophila* Tcf and Groucho interact to repress Wingless signalling activity. *Nature* 395:604–8
12. Roose J, Molenaar M, Peterson J, Hurenkamp J, Brantjes H, et al. 1998. The *Xenopus* Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* 395:608–12

13. van de Wetering M, Cavallo R, Dooijes D, van Beest M, van Es J, et al. 1997. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. *Cell* 88:789–99
14. Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, et al. 2006. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* 124:47–59
15. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, et al. 1997. Constitutive transcriptional activation by a β-catenin-Tcf complex in $APC^{-/-}$ colon carcinoma. *Science* 275:1784–87
16. Sansom OJ, Reed KR, Hayes AJ, Ireland H, Brinkmann H, et al. 2004. Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev.* 18:1385–90
17. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, et al. 2002. The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111:241–50
18. van der Flier LG, Sabates-Bellver J, Oving I, Haegebarth A, De Palo M, et al. 2007. The intestinal Wnt/TCF signature. *Gastroenterology* 132:628–32
19. Gaspar C, Cardoso J, Franken P, Molenaar L, Morreau H, et al. 2008. Cross-species comparison of human and mouse intestinal polyps reveals conserved mechanisms in adenomatous polyposis coli (APC)-driven tumorigenesis. *Am. J. Pathol.* 172:1363–80
20. Stappenbeck TS, Mills JC, Gordon JI. 2003. Molecular features of adult mouse small intestinal epithelial progenitors. *Proc. Natl. Acad. Sci. USA* 100:1004–9
21. Kosinski C, Li VS, Chan AS, Zhang J, Ho C, et al. 2007. Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc. Natl. Acad. Sci. USA* 104:15418–23
22. Yochum GS, McWeeney S, Rajaraman V, Cleland R, Peters S, Goodman RH. 2007. Serial analysis of chromatin occupancy identifies β-catenin target genes in colorectal carcinoma cells. *Proc. Natl. Acad. Sci. USA* 104:3324–29
23. Hatzis P, van der Flier LG, van Driel MA, Guryev V, Nielsen F, et al. 2008. Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol. Cell. Biol.* 28:2732–44
24. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. 2005. Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* 129:626–38
25. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, et al. 1998. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* 19:379–83
26. Fevr T, Robine S, Louvard D, Huelsken J. 2007. Wnt/β-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. *Mol. Cell. Biol.* 27:7551–59
27. Kuhnert F, Davis CR, Wang HT, Chu P, Lee M, et al. 2004. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl. Acad. Sci. USA* 101:266–71
28. Pinto D, Gregorieff A, Begthel H, Clevers H. 2003. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* 17:1709–13
29. Kim KA, Kakitani M, Zhao J, Oshima T, Tang T, et al. 2005. Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* 309:1256–59
30. van Es JH, Jay P, Gregorieff A, van Gijn ME, Jonkheer S, et al. 2005. Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nat. Cell Biol.* 7:381–86
31. Bjerknes M, Cheng H. 2006. Intestinal epithelial stem cells and progenitors. *Methods Enzymol.* 419:337–83
32. Marshman E, Booth C, Potten CS. 2002. The intestinal epithelial stem cell. *Bioessays* 24:91–98
33. Potten CS, Kovacs L, Hamilton E. 1974. Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet.* 7:271–83
34. Potten CS. 1977. Extreme sensitivity of some intestinal crypt cells to X and gamma irradiation. *Nature* 269:518–21
35. Potten CS, Owen G, Booth D. 2002. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* 115:2381–88
36. Bjerknes M, Cheng H. 1981. The stem-cell zone of the small intestinal epithelium. I. Evidence from Paneth cells in the adult mouse. *Am. J. Anat.* 160:51–63
37. Bjerknes M, Cheng H. 1981. The stem-cell zone of the small intestinal epithelium. III. Evidence from columnar, enteroendocrine, and mucous cells in the adult mouse. *Am. J. Anat.* 160:77–91

38. Bjerknes M, Cheng H. 1999. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* 116:7–14
39. Cheng H, Leblond CP. 1974. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian theory of the origin of the four epithelial cell types. *Am. J. Anat.* 141:537–61
40. Cheng H, Leblond CP. 1974. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. *Am. J. Anat.* 141:461–79
41. Nicolas P, Kim KM, Shibata D, Tavare S. 2007. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput. Biol.* 3:e28
42. Yatabe Y, Tavare S, Shibata D. 2001. Investigating stem cells in human colon by using methylation patterns. *Proc. Natl. Acad. Sci. USA* 98:10839–44
43. Cairnie AB, Millen BH. 1975. Fission of crypts in the small intestine of the irradiated mouse. *Cell Tissue Kinet.* 8:189–96
44. Schmidt GH, Winton DJ, Ponder BA. 1988. Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine. *Development* 103:785–90
45. Fuller CE, Davies RP, Williams GT, Williams ED. 1990. Crypt restricted heterogeneity of goblet cell mucus glycoprotein in histologically normal human colonic mucosa: a potential marker of somatic mutation. *Br. J. Cancer* 61:382–84
46. Novelli M, Cossu A, Oukrif D, Quaglia A, Lakhani S, et al. 2003. X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. *Proc. Natl. Acad. Sci. USA* 100:3311–14
47. Novelli MR, Williamson JA, Tomlinson IP, Elia G, Hodgson SV, et al. 1996. Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* 272:1187–90
48. Taylor RW, Barron MJ, Borthwick GM, Gospel A, Chinnery PF, et al. 2003. Mitochondrial DNA mutations in human colonic crypt stem cells. *J. Clin. Investig.* 112:1351–60
49. Winton DJ, Blount MA, Ponder BA. 1988. A clonal marker induced by mutation in mouse intestinal epithelium. *Nature* 333:463–66
50. Winton DJ, Ponder BA. 1990. Stem-cell organization in mouse small intestine. *Proc. Biol. Sci.* 241:13–18
51. Barker N, van Es JH, Kuipers J, Kujala P, Van Den Born M, et al. 2007. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449:1003–7
52. Soriano P. 1999. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nat. Genet.* 21:70–71
53. Schofield R. 1978. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* 4:7–25
54. Mills JC, Gordon JI. 2001. The intestinal stem cell niche: There grows the neighborhood. *Proc. Natl. Acad. Sci. USA* 98:12334–36
55. Powell DW, Mifflin RC, Valentich JD, Crowe SE, Saada JI, West AB. 1999. Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *Am. J. Physiol.* 277:C183–201
56. Haramis AP, Begthel H, Van Den Born M, van Es J, Jonkheer S, et al. 2004. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* 303:1684–86
57. Hardwick JC, Van Den Brink GR, Bleuming SA, Ballester I, Van Den Brande JM, et al. 2004. Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon. *Gastroenterology* 126:111–21
58. Massague J, Seoane J, Wotton D. 2005. Smad transcription factors. *Genes Dev.* 19:2783–810
59. He XC, Zhang J, Tong WG, Tawfik O, Ross J, et al. 2004. BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-β-catenin signaling. *Nat. Genet.* 36:1117–21
60. Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, et al. 2001. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat. Genet.* 28:184–87
61. Howe JR, Roth S, Ringold JC, Summers RW, Järvinen HJ, et al. 1998. Mutations in the *SMAD4/DPC4* gene in juvenile polyposis. *Science* 280:1086–88
62. Zhou XP, Woodford-Richens K, Lehtonen R, Kurose K, Aldred M, et al. 2001. Germline mutations in BMPR1A/ALK3 cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes. *Am. J. Hum. Genet.* 69:704–11

63. Kaestner KH, Silberg DG, Traber PG, Schutz G. 1997. The mesenchymal winged helix transcription factor Fkh6 is required for the control of gastrointestinal proliferation and differentiation. *Genes Dev.* 11:1583–95
64. Pabst O, Zweigerdt R, Arnold HH. 1999. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development* 126:2215–25
65. Wang Y, Wang L, Iordanov H, Swietlicki EA, Zheng Q, et al. 2006. Epimorphin$^{-/-}$ mice have increased intestinal growth, decreased susceptibility to dextran sodium sulfate colitis, and impaired spermatogenesis. *J. Clin. Investig.* 116:1535–46
66. Madison BB, Braunstein K, Kuizon E, Portman K, Qiao XT, Gumucio DL. 2005. Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* 132:279–89
67. Ramalho-Santos M, Melton DA, McMahon AP. 2000. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* 127:2763–72
68. Karlsson L, Lindahl P, Heath JK, Betsholtz C. 2000. Abnormal gastrointestinal development in PDGF-A and PDGFR-α deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis. *Development* 127:3457–66
69. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, et al. 1998. Identification of c-MYC as a target of the APC pathway. *Science* 281:1509–12
70. Muncan V, Sansom OJ, Tertoolen L, Phesse TJ, Begthel H, et al. 2006. Rapid loss of intestinal crypts upon conditional deletion of the Wnt/Tcf-4 target gene c-*Myc*. *Mol. Cell Biol.* 26:8418–26
71. Sansom OJ, Meniel VS, Muncan V, Phesse TJ, Wilkins JA, et al. 2007. Myc deletion rescues Apc deficiency in the small intestine. *Nature* 446:676–79
72. Miller H, Zhang J, Kuolee R, Patel GB, Chen W. 2007. Intestinal M cells: the fallible sentinels? *World J. Gastroenterol.* 13:1477–86
73. Sbarbati A, Osculati F. 2005. A new fate for old cells: brush cells and related elements. *J. Anat.* 206:349–58
74. Artavanis-Tsakonas S, Rand MD, Lake RJ. 1999. Notch signaling: cell fate control and signal integration in development. *Science* 284:770–76
75. Tamura K, Taniguchi Y, Minoguchi S, Sakai T, Tun T, et al. 1995. Physical interaction between a novel domain of the receptor Notch and the transcription factor RBP-Jκ/Su(H). *Curr. Biol.* 5:1416–23
76. Milano J, McKay J, Dagenais C, Foster-Brown L, Pognan F, et al. 2004. Modulation of notch processing by γ-secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol. Sci.* 82:341–58
77. Wong GT, Manfra D, Poulet FM, Zhang Q, Josien H, et al. 2004. Chronic treatment with the γ-secretase inhibitor LY-411,575 inhibits β-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation. *J. Biol. Chem.* 279:12876–82
78. van Es JH, van Gijn ME, Riccio O, Van Den Born M, Vooijs M, et al. 2005. Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* 435:959–63
79. Vooijs M, Ong CT, Hadland B, Huppert S, Liu Z, et al. 2007. Mapping the consequence of Notch1 proteolysis in vivo with NIP-CRE. *Development* 134:535–44
80. Riccio O, van Gijn ME, Bezdek AC, Pellegrinet L, van Es JH, et al. 2008. Loss of intestinal crypt progenitor cells owing to inactivation of both Notch1 and Notch2 is accompanied by derepression of CDK inhibitors p27^{Kip1} and p57^{Kip2}. *EMBO Rep.* 9:377–83
81. Fre S, Huyghe M, Mourikis P, Robine S, Louvard D, Artavanis-Tsakonas S. 2005. Notch signals control the fate of immature progenitor cells in the intestine. *Nature* 435:964–68
82. Stanger BZ, Datar R, Murtaugh LC, Melton DA. 2005. Direct regulation of intestinal fate by Notch. *Proc. Natl. Acad. Sci. USA* 102:12443–48
83. Jensen J, Pedersen EE, Galante P, Hald J, Heller RS, et al. 2000. Control of endodermal endocrine development by Hes-1. *Nat. Genet.* 24:36–44
84. Suzuki K, Fukui H, Kayahara T, Sawada M, Seno H, et al. 2005. Hes1-deficient mice show precocious differentiation of Paneth cells in the small intestine. *Biochem. Biophys. Res. Commun.* 328:348–52
85. Yang Q, Bermingham NA, Finegold MJ, Zoghbi HY. 2001. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* 294:2155–58

86. Shroyer NF, Wallis D, Venken KJ, Bellen HJ, Zoghbi HY. 2005. Gfi1 functions downstream of Math1 to control intestinal secretory cell subtype allocation and differentiation. *Genes Dev.* 19:2412–17
87. Amann JM, Chyla BJ, Ellis TC, Martinez A, Moore AC, et al. 2005. Mtgr1 is a transcriptional corepressor that is required for maintenance of the secretory cell lineage in the small intestine. *Mol. Cell. Biol.* 25:9576–85
88. Karam SM. 1999. Lineage commitment and maturation of epithelial cells in the gut. *Front Biosci.* 4:D286–98
89. Crosnier C, Vargesson N, Gschmeissner S, Ariza-McNaughton L, Morrison A, Lewis J. 2005. Delta-Notch signalling controls commitment to a secretory fate in the zebrafish intestine. *Development* 132:1093–104
90. Katz JP, Perreault N, Goldstein BG, Lee CS, Labosky PA, et al. 2002. The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* 129:2619–28
91. Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, et al. 2002. Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* 295:1726–29
92. Schonhoff SE, Giel-Moloney M, Leiter AB. 2004. Minireview: development and differentiation of gut endocrine cells. *Endocrinology* 145:2639–44
93. Jenny M, Uhl C, Roche C, Duluc I, Guillermin V, et al. 2002. Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *EMBO J.* 21:6338–47
94. Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, et al. 1997. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev.* 11:2323–34
95. Desai S, Loomis Z, Pugh-Bernard A, Schrunk J, Doyle MJ, et al. 2008. Nkx2.2 regulates cell fate choice in the enteroendocrine cell lineages of the intestine. *Dev. Biol.* 313:58–66
96. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, et al. 1996. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 122:983–95
97. Hill ME, Asa SL, Drucker DJ. 1999. Essential requirement for Pax6 in control of enteroendocrine proglucagon gene transcription. *Mol. Endocrinol.* 13:1474–86
98. Larsson LI, St-Onge L, Hougaard DM, Sosa-Pineda B, Gruss P. 1998. Pax 4 and 6 regulate gastrointestinal endocrine cell development. *Mech. Dev.* 79:153–59
99. Batlle E, Henderson JT, Beghtel H, Van Den Born MM, Sancho E, et al. 2002. β-Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 111:251–63
100. Cortina C, Palomo-Ponce S, Iglesias M, Fernandez-Masip JL, Vivancos A, et al. 2007. EphB-ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells. *Nat. Genet.* 39:1376–83
101. Blache P, van de Wetering M, Duluc I, Domon C, Berta P, et al. 2004. SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the *CDX2* and *MUC2* genes. *J. Cell Biol.* 166:37–47
102. Bastide P, Darido C, Pannequin J, Kist R, Robine S, et al. 2007. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *J. Cell Biol.* 178:635–48
103. Mori-Akiyama Y, Van Den Born M, van Es JH, Hamilton SR, Adams HP, et al. 2007. SOX9 is required for the differentiation of Paneth cells in the intestinal epithelium. *Gastroenterology* 133:539–46
104. Ng AY, Waring P, Ristevski S, Wang C, Wilson T, et al. 2002. Inactivation of the transcription factor Elf3 in mice results in dysmorphogenesis and altered differentiation of intestinal epithelium. *Gastroenterology* 122:1455–66
105. Flentjar N, Chu PY, Ng AY, Johnstone CN, Heath JK, et al. 2007. TGF-βRII rescues development of small intestinal epithelial cells in Elf3-deficient mice. *Gastroenterology* 132:1410–19
106. Haegebarth A, Bie W, Yang R, Crawford SE, Vasioukhin V, et al. 2006. Protein tyrosine kinase 6 negatively regulates growth and promotes enterocyte differentiation in the small intestine. *Mol. Cell. Biol.* 26:4949–57
107. Fodde R, Brabletz T. 2007. Wnt/β-catenin signaling in cancer stemness and malignant behavior. *Curr. Opin. Cell Biol.* 19:150–58

108. Miyaki M, Konishi M, Kikuchi-Yanoshita R, Enomoto M, Igari T, et al. 1994. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. *Cancer Res.* 54:3011–20
109. Miyoshi Y, Ando H, Nagase H, Nishisho I, Horii A, et al. 1992. Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc. Natl. Acad. Sci. USA* 89:4452–56
110. Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, et al. 1992. APC mutations occur early during colorectal tumorigenesis. *Nature* 359:235–37
111. Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, et al. 2000. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. *Nat. Genet.* 26:146–47
112. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, et al. 1997. Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC. *Science* 275:1787–90
113. Rubinfeld B, Robbins P, El-Gamil M, Albert I, Porfiri E, Polakis P. 1997. Stabilization of β-catenin by genetic defects in melanoma cell lines. *Science* 275:1790–92
114. Barker N, Huls G, Korinek V, Clevers H. 1999. Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium. *Am. J. Pathol.* 154:29–35
115. Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. *Cell* 61:759–67
116. Moser AR, Pitot HC, Dove WF. 1990. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* 247:322–24
117. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, et al. 1992. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* 256:668–70
118. Gaspar C, Fodde R. 2004. APC dosage effects in tumorigenesis and stem cell differentiation. *Int. J. Dev. Biol.* 48:377–86
119. Andreu P, Colnot S, Godard C, Gad S, Chafey P, et al. 2005. Crypt-restricted proliferation and commitment to the Paneth cell lineage following Apc loss in the mouse intestine. *Development* 132:1443–51
120. Shibata H, Toyama K, Shioya H, Ito M, Hirota M, et al. 1997. Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. *Science* 278:120–23
121. Harada N, Tamai Y, Ishikawa T, Sauer B, Takaku K, et al. 1999. Intestinal polyposis in mice with a dominant stable mutation of the β-catenin gene. *EMBO J.* 18:5931–42
122. Dick JE. 2003. Breast cancer stem cells revealed. *Proc. Natl. Acad. Sci. USA* 100:3547–49
123. Reya T, Morrison SJ, Clarke MF, Weissman IL. 2001. Stem cells, cancer, and cancer stem cells. *Nature* 414:105–11
124. Wang JC, Dick JE. 2005. Cancer stem cells: lessons from leukemia. *Trends Cell Biol.* 15:494–501
125. O'Brien CA, Pollett A, Gallinger S, Dick JE. 2007. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445:106–10
126. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, et al. 2007. Identification and expansion of human colon-cancer-initiating cells. *Nature* 445:111–15
127. Kelly PN, Dakic A, Adams JM, Nutt SL, Strasser A. 2007. Tumor growth need not be driven by rare cancer stem cells. *Science* 317:337

Outline of this thesis

As described in the introduction, the intestinal epithelium is a rapidly self-renewing tissue. During the last decade, it has been shown that the Wnt pathway is the primary driving force behind this proliferation. The initiating objective of this thesis was to investigate which genes are regulated by the Wnt/TCF pathway in the intestine and in colorectal cancer (CRC). In chapters 2 and 3, the identification of intestinal Wnt target genes is described based on expression array analysis of CRC cell lines and adenoma and adenocarcinoma samples from human patients. To investigate which of these target genes are directly regulated by the Wnt pathway, we used a genome-wide analysis of TCF4-associated chromatin in LS174T CRC cells (chapter 4). One of the identified Wnt target genes was the transcription factor *Ascl2*. In the intestine this gene has a stem cell-restricted expression pattern. In chapter 5, the role of Ascl2 in controlling intestinal stem cell fate is described, based on both gain- and loss-of-function experiments. Chapter 6 describes two mouse models that will be used for the further characterization of intestinal stem cells. Chapter 7 describes the generation of three mouse models that will be used to study the function of the *Regenerating* (*Reg*) gene family in the intestine. In chapter 8 the different chapters are summarized and discussed.

Chapter 2

The Intestinal Wnt/TCF Signature

Reprinted from: Gastroenterology. 2007 Feb;132(2):628-32

The Intestinal Wnt/TCF Signature

LAURENS G. VAN DER FLIER,* JACOB SABATES–BELLVER,‡ IRMA OVING,* ANDREA HAEGEBARTH,* MARIAGRAZIA DE PALO,§ MARCELLO ANTI,§ MARIELLE E. VAN GIJN,* SASKIA SUIJKERBUIJK,* MARC VAN DE WETERING,* GIANCARLO MARRA,‡ and HANS CLEVERS*

*Hubrecht Institute, Netherlands Institute for Developmental Biology and Centre for Biomedical Genetics, Utrecht, The Netherlands; ‡Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland; and §the Gastroenterology Unit, Belcolle Hospital, Strada Sammartinese, Viterbo, Italy

Background & Aims: **In colorectal cancer, activating mutations in the Wnt pathway transform epithelial cells through the inappropriate expression of a TCF4 target gene program, which is physiologically expressed in intestinal crypts.** ***Methods:*** **We have now performed an exhaustive array-based analysis of this target gene program in colorectal cancer cell lines carrying an inducible block of the Wnt cascade. Independently, differential gene-expression profiles of human adenomas and adenocarcinomas vs normal colonic epithelium were obtained.** ***Results:*** **Expression analyses of approximately 80 genes common between these data sets were performed in a murine adenoma model. The combined data sets describe a core target gene program, the intestinal Wnt/TCF signature gene set, which is responsible for the transformation of human intestinal epithelial cells.** ***Conclusions:*** **The genes were invariably expressed in adenomas, yet could be subdivided into 3 modules, based on expression in distinct crypt compartments. A module of 17 genes was specifically expressed at the position of the crypt stem cell.**

The overwhelming majority of colorectal cancers (CRCs) is initiated by activating mutations in the Wnt pathway.[1,2] These mutations either remove the tumor suppressors adenomatous polyposis coli (APC), or axin or activate the proto-oncogene β-catenin. As a common result, β-catenin accumulates in the nucleus, and constitutively binds to the T-cell factor (TCF)4 transcription factor, resulting in transcriptional activation of Wnt/TCF4 target genes, initiating transformation of intestinal epithelial cells.[3,4] Physiologically, the Wnt pathway is essential for the maintenance of crypt progenitor compartments, as evidenced by mice lacking the TCF4 transcription factor,[5] or transgenically expressing the secreted Dickkopf-1 Wnt inhibitor.[6,7]

A multitude of reports have appeared in which single-candidate TCF4 target genes have been described (eg, references listed in Table 1 and available at http://www.stanford.edu/~rnusse/pathways/targets.html). Typically, these studies describe differential expression between cells with and without an activated Wnt pathway, followed by transient promoter assays. DNA array technology allows the assessment of differential messenger RNA (mRNA) expression on a genome-wide scale.

Materials and Methods

Cell Culture

CRC cell lines LS174T and DLD1, stably expressing inducible dominant-negative (dn)TCF1 or dnTCF4, were generated as previously described.[8] The Wnt pathway activity in the CRC cells was determined as described previously[3] using the optimized TCF reporter pTopGlow and its negative control pFopGlow, constructed in our laboratory.[9]

Oligonucleotide Microarray Analysis of CRC Cell Lines

RNA was isolated after 10 and 20 hours induction of the dnTCFs. RNA quality was assessed using capillary gel electrophoresis (BioAnalyzer; Agilent Technologies). Complementary DNA (cDNA) synthesis and labeling was performed according to Affymetrix (Santa Clara, CA) guidelines. cRNA was synthesized and labeled with Affymetrix One-cycle Target Labeling kit, and hybridized on Affymetrix GeneChip HG-U133 plus 2.0 microarrays. The overall fluorescence for each GeneChip was scaled to a target intensity of 200. The expression profiles at 10 and 20 hours after induction were compared with those of noninduced controls by pair-wise comparisons performed with GeneChip Operating Software (Affymetrix). Only probes with a significantly decreased call for both time points were included.

Oligonucleotide Microarray Analysis of Human Tissues

Colorectal adenomas were collected during endoscopy at a single gastroenterology unit (Belcolle Hospital, Viterbo, Italy) with full institutional review board approval. In each patient, normal mucosa also was collected at about a 2–4-cm distance from the adenoma. Biopsy specimens

Abbreviations used in this paper: APC, adenomatous polyposis coli; CRC, colorectal cancer; dn, dominant-negative; TCF, T-cell factor; wt, wild type.

Table 1. Target Genes Down-Regulated in All 4 CRC Cell Lines on Over Expression of dnTCFs

			LS174T dnTCF1		LS174T dnTCF4		DLD1 dnTCF1		DLD1	
Gene symbol	References	Affymetrix ID	10 h	20 h	10 h	20 h	10 h	20 h	10 h	20 h
ASCL2	[18]	229215_at	−4.3	−24	−2.3	−15	−2.8	−7.5	−3.0	−4.0
AXIN2	[14–16]	222696_at	−2.5	−4.9	−2.3	−2.6	−2.8	−3.5	−2.8	−3.7
BMP4[a]	[21]	211518_s_at	−2.0	−2.5	−2.3	−4.9	−3.0	−3.2	−1.6	−2.3
C1orf33[a]		220688_s_at	−1.7	−3.2	−2.3	−2.8	−1.7	−2.0	−1.6	−2.0
HIG2	[20]	1554452_a_at	−1.7	−4.0	−2.1	−3.2	−1.4	−2.6	−1.9	−2.0
HSPC111		203023_at	−1.5	−3.0	−2.3	−2.5	−1.4	−1.6	−1.9	−1.7
HSPC111		214011_s_at	−1.6	−3.0	−2.3	−2.3	−1.5	−1.9	−1.6	−1.6
KITLG[a]		226534_at	−2.6	−3.0	−2.3	−2.3	−4.3	−2.6	−1.6	−2.8
LGR5[a]	[19]	213880_at	−4.3	−9.8	−2.3	−3.5	−7.5	−7.5	−2.1	−3.2
MYC[a]	[17]	202431_s_at	−2.1	−2.3	−2.3	−1.6	−2.8	−3.0	−2.1	−2.3
NOL1		214427_at	−1.7	−2.8	−2.3	−2.1	−1.3	−1.9	−1.6	−1.6
PPIF		201490_s_at	−1.4	−1.6	−2.3	−2.0	−1.5	−1.9	−1.5	−1.5
SOX4	[22]	201416_at	−1.3	−2.1	−2.3	−2.1	−3.0	−3.2	−2.1	−2.0
WDR71		218957_s_at	−2.8	−17	−2.3	−2.8	−1.9	−3.7	−2.0	−3.7
ZIC2		223642_at	−1.7	−2.3	−2.3	−2.3	−1.4	−2.1	−1.7	−1.9
ZNRF3[a]		226360_at	−4.6	−3.2	−2.3	−2.5	−3.5	−3.2	−2.1	−2.8

NOTE. Fold changes of genes down-regulated in all 4 CRC cell lines on over expression of dnTCFs are shown.
[a]Genes also have been identified in the Stanford array experiment.[8]

were immersed immediately in RNAlater (Ambion, Huntingdon, UK), homogenized, and RNA was extracted with RNeasy kit (QIAgen, Basel, Switzerland). RNA quality was verified by capillary gel electrophoresis and a total of 32 pairs of normal mucosa and adenomas was analyzed. cRNA was synthesized, labeled, and hybridized as described earlier. GeneSpring software (Silicon Genetics, Redwood City, CA) was used for gene expression data and statistical analyses. The nonparametric Mann–Whitney test was used with a false discovery rate of .05 and Bonferroni correction for the group comparison analysis (normal mucosa vs adenomas or cancer). Transcriptome data from a series of 25 previously collected colon adenocarcinomas and 10 samples of normal mucosa from some of these patients were obtained with the same procedure[10] and used in this study.

In Situ Hybridizations

Mouse orthologs of selected transcripts were obtained as expressed sequence tags IMAGE consortium (MRC geneservise, Badrahman, UK) or RZPD (German Resource Center for Genome Research, Berlin, Germany). These clones were used for in vitro transcription reactions to generate probes for in situ hybridizations. Protocols for in vitro transcription and in situ hybridizations are described elsewhere.[11]

Results

Inhibition of the Constitutively Active Wnt Pathway in CRC Cells

N-terminally truncated TCFs do not bind β-catenin and act as potent inhibitors of endogenous β-catenin/TCF complexes.[12] Tcf4 is expressed physiologically in the intestine.[13] In our hands, the DNA-binding characteristics of TCF4 and TCF1 are essentially identical. For TCF target gene identification, we generated a panel of cell clones from the CRC cell lines LS174T (mutationally activated allele of the *CTNNB1* gene encoding β-catenin) and in DLD1 (mutant APC), in which the Wnt cascade could be inhibited by inducible expression of either dnTCF1 or dnTCF4. Individ-

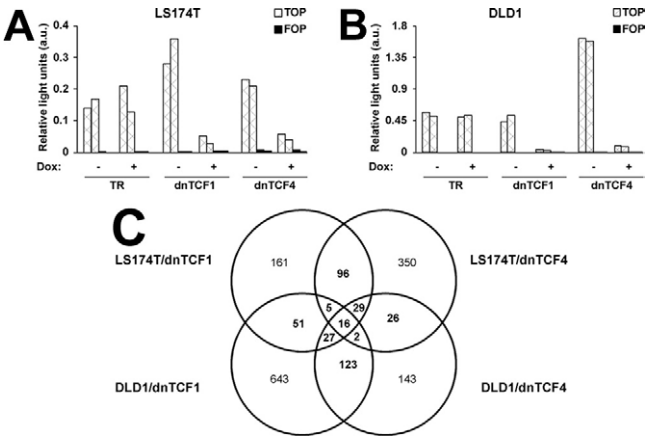


Figure 1. Over expression of dnTCFs inhibits the Wnt pathway in CRC cells. TCF/β-catenin driven transcription is abrogated by over expression of dnTCF1 or dnTCF4 in (*A*) LS174T cells and (*B*) DLD1 cells. The activity of the TCF reporter, pTopGlow (□, TOP), and the control, pFopGlow (■, FOP), after 20 hours with or without doxycycline treatment is shown. Parental cells that expressed the tetracycline repressor (TR) were used as controls. Renilla luciferase levels were used as transfection controls. (*C*) Venn diagram showing the (overlapping) down-regulated probes on dnTCF1 or dnTCF4 induction. Selection is based on down-regulation in 2 or more of the 4 cell-line transfectants used in this study. Not shown in the figure is the overlap in 6 unique probes between LS174T/dnTCF1 and DLD1/dnTCF4 and 6 probes between LS174T/dnTCF4 and DLD1/dnTCF1.

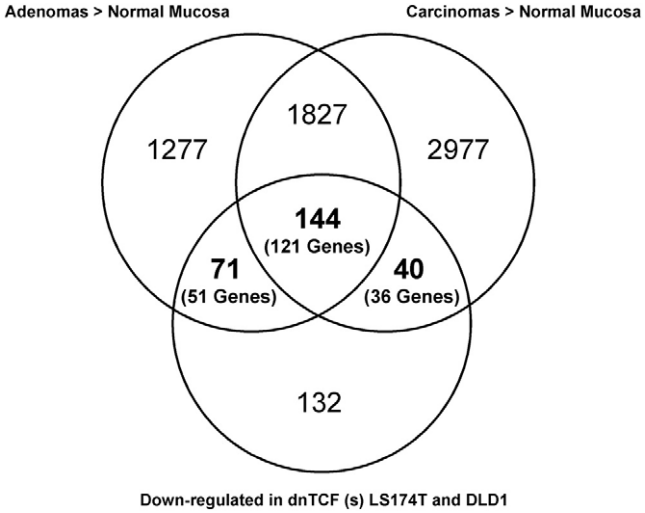


Figure 2. The intestinal Wnt/TCF signature gene set. TCF/β-catenin target genes were selected by comparing the 387 probes (lower circle) obtained with the procedure described in Figure 1*C* with the 8307 probes whose expression levels in human tumors were increased significantly, relative to those in normal mucosa. The left upper circle represents 3319 probes increased in adenomas, whereas the right upper circle represents 4988 probes increased in carcinomas (1971 were increased in both tumor types). With this approach, we classified 51 genes (71 probes) up-regulated in adenomas, 36 genes (40 probes) up-regulated in carcinomas, and 121 genes (144 probes) up-regulated in both tumor types as β-catenin/TCF target genes. These gene lists are reported in Supplementary Table 1 (supplementary material online at www.gastrojournal.org).

ual transfectants were selected based on the induced inhibition of the constitutive Wnt activity. For the 4 selected clones, the constitutively active Wnt pathway could be inhibited close to background levels on doxcycycline-mediated induction of dnTCF1 or dnTCF4 (Figure 1*A* and *B*). Moreover, all clones underwent a rapid, robust G1 arrest within 24 hours (not shown).

The Genetic Program Driven by β-Catenin/TCF in CRC Cell Lines

We originally subjected the LS174T/dnTCF4 clone to a microarray experiment on the Stanford spotted-cDNA array platform containing 24,000 probes.[8] To obtain a more comprehensive list of TCF target genes in intestinal cancer cells, we performed expression profiling using the Affymetrix GeneChip HG-U133 plus 2.0, which contains 54,675 probes. We initiated analyses with the LS174T/dnTCF4 cells. mRNA was isolated at 10 and 20 hours after induction and from noninduced control cells. The 115 probes reported on the Stanford platform[8] represented 101 significantly down-regulated genes, all of which also were present on the Affymetrix arrays. More than 60% of these 101 genes turned out to be down-regulated significantly at either the 10- or 20-hour time points in the Affymetrix measurements, whereas 26% were significantly down-regulated at both time points. This showed good reproducibility and robustness of the assay, given the extensive differences in technology between the 2 platforms.

Similar array experiments then were performed for the other 3 transfectants (LS174T/dnTCF1, DLD1/dnTCF4, and DLD1/dnTCF1). For each cell line, we compiled lists of probe features significantly down-regulated at both time points. Figure 1*C* gives a Venn diagram summarizing the down-regulated features shared between the 4 different cell

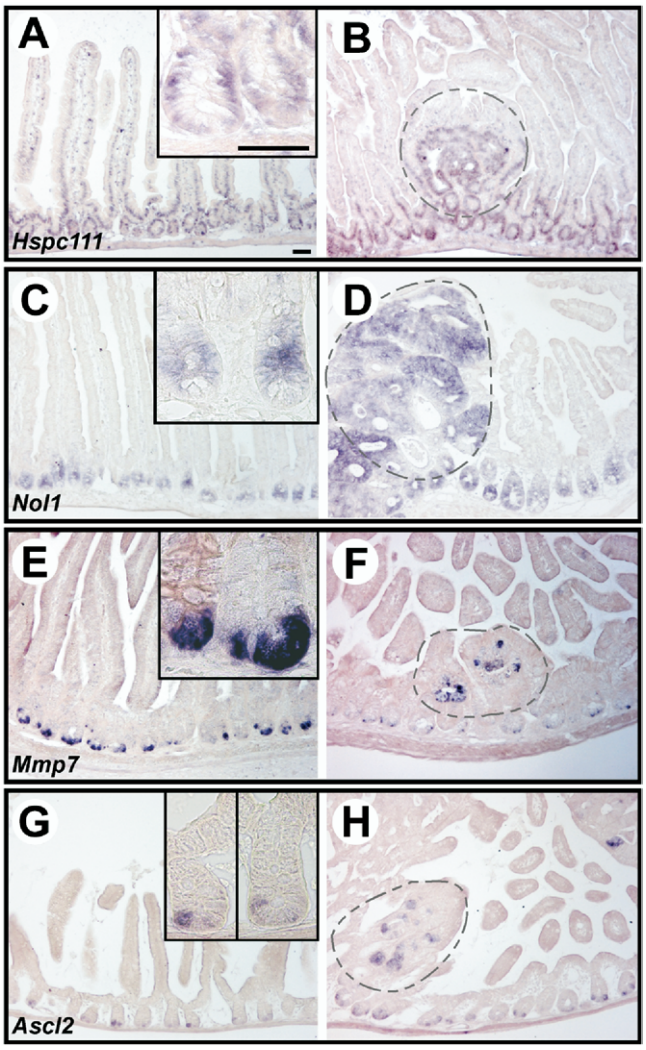


Figure 3. Confirmation and classification of TCF target genes by in situ hybridization on intestines of adult *APC*^{min} mice. All Wnt signature genes tested were expressed in (*A*, *C*, *E*, and *G*) wild type crypts and in (*B*, *D*, *F*, and *H*) adenomas of *Apc*^{min} mice. The staining patterns in wt crypts could be divided into 3 categories. (*A* and *B*) An example of the expression of a gene within the first category, the mouse ortholog of *HSPC111*, is shown. In addition to expression within the proliferative compartment of crypts and adenomas, a decreasing gradient of signal is observed along the base of the villus. (*C* and *D*) Another gene in this category, the mouse ortholog for *NOL1*, showed expression restricted to the proliferative compartment of crypts and adenomas, but lacked significant expression along the villus. (*E* and *F*) The second staining pattern involves the Paneth cells at the bottom of the crypts, as represented by the mouse ortholog for *MMP7*. (*G* and *H*) In situ hybridizations for the mouse ortholog of *ASCL2* defined a third category that revealed expression in adenomas and in a few cells near the crypt bottom. This position coincides with the location of the elusive crypt stem cells. Size bars represent 50 μm, and adenomas are encircled by dotted lines.

lines. The 15 genes down-regulated in all 4 cell lines are reported in Table 1. Table 1 includes some well-known Wnt targets *AXIN2*,[14–16] *c-Myc*,[17] *ASCL2*,[18] *LGR5*,[19] *HIG2*,[20] *BMP4*,[21] and *SOX4*,[22] as well as 8 novel TCF target genes.

To obtain an information-rich, yet robust data set, we arbitrarily selected the probes that were down-regulated at both time points in at least 2 different cell lines. This resulted in a list that contained 387 probe features. This list was used for comparison with the gene-expression profiles of adenomas and adenocarcinomas described later.

Gene Expression Profiles of Human Adenomas and Adenocarcinomas

The transcriptome of 32 adenomas and 25 carcinomas was analyzed and compared with that of normal colonic mucosa. The genes up-regulated in adenomas and/or carcinomas were compared with those of the 387 β-catenin/TCF responsive genes identified in the dnTCF CRC cell lines. Figure 2 shows the Venn diagram procedure we used to select the Wnt/TCF signature gene set. A total of 255 common probes corresponding to 208 different genes are reported in Supplemental Table 1 (supplementary material online at www.gastrojournal.org). A total of 121 of the genes were up-regulated in both tumor types, whereas 51 and 36 were up-regulated only in the adenomas or in carcinomas, respectively.

Confirmation of Target Genes by In Situ Hybridizations

In our initial study,[8] we observed that, as a rule, the identified TCF target genes were expressed physiologically in rapidly dividing crypt cells. It since has been observed that a subset of TCF target genes is expressed in the postmitotic Paneth cells, which are located at the crypt bottom.[23,24] We extended these observations for the intestinal Wnt/TCF signature gene set by in situ hybridizations on intestinal tissue derived from adult mice carrying the Apc^{min} allele.[25] In situ hybridizations were performed for approximately 80 genes (Figure 3 and Supplemental Table 2; supplementary material online at www.gastrojournal.org). All tested genes were expressed in the Apc^{min} adenomas. The staining patterns in crypts could be grouped in several categories.

The first category consisted of about 80% of the tested genes. These genes were expressed in the proliferative compartment of the crypts. Examples are given in Figure 3*A–D*.

The second category of TCF target genes comprised the Paneth cell maturation markers.[23,24] The tyrosine kinase receptor EphB3, present in our Wnt/TCF signature gene set, falls in this category.[26] Other Paneth cell–specific genes such as *MMP7* (Figure 3*E* and *F*) and defensin-6 are not represented in the signature because they were not expressed in the cell lines. They were, however, clearly up-regulated in the tumor samples in the current study.

A third category (17 genes; Supplementary Table 2) yielded staining in 1–5 crypt cells, typically located near the crypt bottom. The cells were distinct from the Paneth cells. The location was highly reminiscent of the position to which the elusive crypt stem cells have been mapped.[27,28] As an example, the expression of the mouse ortholog of *ASCL2* is given in Figure 3*G* and *H*. Of note, a previously published intestinal stem cell marker, *Musashi*,[29,30] in our hands (data not shown) would be best classified in our first category.

Discussion

The current study builds on previous cell line–based work from our laboratory.[8] Here, we provide a comprehensive identification of TCF4 targets in 2 different cell lines carrying 2 different dnTCF genes, and by performing differential gene-expression analysis on a genome-wide oligonucleotide array platform. Moreover, we relate these findings to expression profiles of a set of human adenomas and adenocarcinomas. The Wnt/TCF signature gene set defines the core program activated by TCF4 in intestinal epithelial cells. Because the CRC cell lines used in this study arrest in the G1 phase of the cell cycle on inhibition of the Wnt cascade, this program is essential for the proliferative capacity of CRC cell lines in culture. These observations can be extrapolated to human intestinal tumors, in that the shared TCF4 target gene program likely represents the primary driver behind the transformation behavior of these transformed lesions. Moreover, the individual genes within the signature represent promising targets for therapy of CRC because their expression consistently is activated as the direct result of oncogenic Wnt pathway mutations, although many target genes will be involved causally in the transformed behavior of the neoplastic cells.

The TCF4 target gene program as activated in colorectal neoplasia consists of at least 3 distinct modules, as revealed by studies of physiologic gene expression in murine small intestinal crypts. Genes within one of the modules are expressed by the rapidly dividing crypt progenitors of the transit-amplifying compartment, and are likely the driving force behind the proliferative activity of intestinal neoplasia and cell lines derived thereof. Another module is correlated with maturation of postmitotic Paneth cells, as previously described.[24] We believe that the expression of the Paneth cell module in intestinal neoplasia is fortuitous and does not contribute to malignant transformation.

The discovery of a module expressed at the stem cell position was unexpected. Genes within this module may serve as starting points to study intestinal stem cell biology. A role for Wnt signaling in the biology of the transient-amplifying compartment of crypts has been firmly established previously.[5–7] A similar role in the biology of the intestinal stem cell has so far remained speculative.[31] The TCF target genes in the stem cell module may solidly link physiologic Wnt signaling to intestinal stem cell biology.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1053/j.gastro.2006.08.039.

References

1. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159–170.
2. Nusse R. Wnt signaling in disease and in development. Cell Res 2005;15:28–32.
3. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 1997;275:1784–1787.
4. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 1997;275:1787–1790.
5. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 1998;19:379–383.
6. Kuhnert F, Davis CR, Wang HT, Chu P, Lee M, Yuan J, Nusse R, Kuo CJ. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A 2004;101:266–271.
7. Pinto D, Gregorieff A, Begthel H, Clevers H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev 2003;17:1709–1713.
8. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–250.
9. van de Wetering M, Barker N, Harkes IC, van der Heyden M, Dijk NJ, Hollestelle A, Klijn JG, Clevers H, Schutte M. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61:278–284.
10. di Pietro M, Bellver JS, Menigatti M, Bannwart F, Schnider A, Russell A, Truninger K, Jiricny J, Marra G. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–1059.
11. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 2005;129:626–638.
12. Molenaar M, van de Wetering M, Oosterwegel M, Peterson-Maduro J, Godsave S, Korinek V, Roose J, Destree O, Clevers H. XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell 1996;86:391–399.
13. Barker N, Huls G, Korinek V, Clevers H. Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium. Am J Pathol 1999;154:29–35.
14. Jho EH, Zhang T, Domon C, Joo CK, Freund JN, Costantini F. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol Cell Biol 2002;22:1172–1183.
15. Lustig B, Jerchow B, Sachs M, Weiler S, Pietsch T, Karsten U, van de Wetering M, Clevers H, Schlag PM, Birchmeier W, Behrens J. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Mol Cell Biol 2002;22:1184–1193.
16. Yan D, Wiesmann M, Rohan M, Chan V, Jefferson AB, Guo L, Sakamoto D, Caothien RH, Fuller JH, Reinhard C, Garcia PD, Randazzo FM, Escobedo J, Fantl WJ, Williams LT. Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta-catenin signaling is activated in human colon tumors. Proc Natl Acad Sci U S A 2001;98:14973–14978.
17. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW. Identification of c-MYC as a target of the APC pathway. Science 1998;281:1509–1512.
18. Jubb AM, Chalasani S, Frantz GD, Smits R, Grabsch HI, Kavi V, Maughan NJ, Hillan KJ, Quirke P, Koeppen H. Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. Oncogene 2006;25:3445–3457.
19. Yamamoto Y, Sakamoto M, Fujii G, Tsuiji H, Kenetaka K, Asaka M, Hirohashi S. Overexpression of orphan G-protein-coupled receptor, Gpr49, in human hepatocellular carcinomas with beta-catenin mutations. Hepatology 2003;37:528–533.
20. Kenny PA, Enver T, Ashworth A. Receptor and secreted targets of Wnt-1/beta-catenin signalling in mouse mammary epithelial cells. BMC Cancer 2005;5:3.
21. Kim JS, Crooks H, Dracheva T, Nishanian TG, Singh B, Jen J, Waldman T. Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. Cancer Res 2002;62:2744–2748.
22. Reichling T, Goss KH, Carson DJ, Holdcraft RW, Ley-Ebert C, Witte D, Aronow BJ, Groden J. Transcriptional profiles of intestinal tumors in Apc(Min) mice are unique from those of embryonic intestine and identify novel gene targets dysregulated in human colorectal tumors. Cancer Res 2005;65:166–176.
23. Andreu P, Colnot S, Godard C, Gad S, Chafey P, Niwa-Kawakita M, Laurent-Puig P, Kahn A, Robine S, Perret C, Romagnolo B. Crypt-restricted proliferation and commitment to the Paneth cell lineage following Apc loss in the mouse intestine. Development 2005;132:1443–1451.
24. van Es JH, Jay P, Gregorieff A, van Gijn ME, Jonkheer S, Hatzis P, Thiele A, van den Born M, Begthel H, Brabletz T, Taketo MM, Clevers H. Wnt signalling induces maturation of Paneth cells in intestinal crypts. Nat Cell Biol 2005;7:381–386.
25. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science 1992;256:668–670.
26. Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T, Clevers H. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell 2002;111:251–263.
27. Bjerknes M, Cheng H. Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology 1999;116:7–14.
28. Potten CS. Extreme sensitivity of some intestinal crypt cells to X and gamma irradiation. Nature 1977;269:518–521.
29. Kayahara T, Sawada M, Takaishi S, Fukui H, Seno H, Fukuzawa H, Suzuki K, Hiai H, Kageyama R, Okano H, Chiba T. Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. FEBS Lett 2003;535:131–135.
30. Potten CS, Booth C, Tudor GL, Booth D, Brady G, Hurley P, Ashton G, Clarke R, Sakakibara S, Okano H. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. Differentiation 2003;71:28–41.
31. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature 2005;434:843–850.

Received June 6, 2006. Accepted July 19, 2006.

Address requests for reprints to: Hans Clevers, Hubrecht Institute, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584CT, Utrecht, The Netherlands. e-mail: clevers@niob.knaw.nl; fax: (31) 30-2121-801.

L.G.V.D.F. and J.S.-B. contributed equally to this article.

Chapter 3

Transcriptome Profile of Human Colorectal Adenomas

Reprinted from: Mol Cancer Res. 2007 Dec;5(12):1263-75

Transcriptome Profile of Human Colorectal Adenomas

Jacob Sabates-Bellver,[1] Laurens G. Van der Flier,[4] Mariagrazia de Palo,[5] Elisa Cattaneo,[1] Caroline Maake,[2] Hubert Rehrauer,[3] Endre Laczko,[3] Michal A. Kurowski,[9] Janusz M. Bujnicki,[9] Mirco Menigatti,[7] Judith Luz,[8] Teresa V. Ranalli,[6] Vito Gomes,[6] Alfredo Pastorelli,[5] Roberto Faggiani,[5] Marcello Anti,[5] Josef Jiricny,[1] Hans Clevers,[4] and Giancarlo Marra[1]

[1]Institute of Molecular Cancer Research, [2]Institute of Anatomy, and [3]Functional Genomic Center, University of Zurich, Zurich, Switzerland; [4]Hubrecht Institute, Netherlands Institute for Developmental Biology, Utrecht, the Netherlands; [5]Gastroenterology Unit and [6]Pathology Department, Belcolle Hospital, Viterbo, Italy; [7]Institute of Biochemistry and Genetics and [8]Division of Medical Genetics, University of Basel, Basel, Switzerland; and [9]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland

Abstract

Colorectal cancers are believed to arise predominantly from adenomas. Although these precancerous lesions have been subjected to extensive clinical, pathologic, and molecular analyses, little is currently known about the global gene expression changes accompanying their formation. To characterize the molecular processes underlying the transformation of normal colonic epithelium, we compared the transcriptomes of 32 prospectively collected adenomas with those of normal mucosa from the same individuals. Important differences emerged not only between the expression profiles of normal and adenomatous tissues but also between those of small and large adenomas. A key feature of the transformation process was the remodeling of the Wnt pathway reflected in patent overexpression and underexpression of 78 known components of this signaling cascade. The expression of 19 Wnt targets was closely correlated with clear up-regulation of *KIAA1199*, whose function is currently unknown. In normal mucosa, KIAA1199 expression was confined to cells in the lower portion of intestinal crypts, where Wnt signaling is physiologically active, but it was markedly increased in all adenomas, where it was expressed in most of the epithelial cells, and in colon cancer cell lines, it was markedly reduced by inactivation of the β-catenin/T-cell factor(s) transcription complex, the pivotal mediator of Wnt signaling. Our transcriptomic profiles of normal colonic mucosa and colorectal adenomas shed new light on the early stages of colorectal tumorigenesis and identified *KIAA1199* as a novel target of the Wnt signaling pathway and a putative marker of colorectal adenomatous transformation.

(Mol Cancer Res 2007;5(12):1263–75)

Introduction

In developed countries, sporadic adenomatous colorectal polyps are found in roughly one third of asymptomatic adults below the age of 50 who undergo colonoscopy. Depending on their characteristics (multiplicity, size, histologic features, and degree of dysplasia), these lesions can be associated with a substantial risk of recurrence (up to 60% at 3 years) and the development of advanced neoplastic disease (reviewed in ref. 1 and references therein). It has been estimated that 15% of all adenomas measuring ≥1 cm will progress to carcinomas within 10 years of their detection (2).

Although adenomatous polyps are not the only precancerous lesions in the colorectum, they are the most common, and they are the precursors of most of the cancers in this organ. In these neoplasms, the transformation process begins in the epithelial crypts and seems to result from qualitative, quantitative, and spatial subversion of the Wnt signaling pathway, the physiologic regulator of epithelial homeostasis (3-5). This adenoma-carcinoma pathway of tumorigenesis is characterized by mutations involving various components of this pathway (e.g., *APC*, whose germ-line mutations are responsible for familial adenomatous polyposis; *CTNNB1*, which encodes a subunit of the cadherin protein complex known as β-catenin; and *Axin*, the gene encoding a multidomain scaffold protein that is essential for β-catenin degradation). The result of these mutations is an accumulation of β-catenin, first in the cytoplasm and then in the nucleus, where it associates with DNA-binding proteins of the T-cell factor (TCF)/lymphoid enhancer factor family, transforming them from transcriptional repressors into transcriptional activators that affect the expression of numerous genes involved in epithelial homeostasis.

Although the key role played by adenomatous polyps in colorectal tumorigenesis is widely acknowledged, the gene expression changes that trigger or accompany their development

Received 6/9/07; revised 7/28/07; accepted 8/2/07.
Grant support: Zurich Cancer League and Swiss National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).
J. Sabates-Bellver and L.G. Van der Flier contributed equally to this work.
Requests for reprints: Giancarlo Marra, Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Phone: 41-044-635-3472; Fax: 41-044-635-3484. E-mail: marra@imcr.uzh.ch

doi:10.1158/1541-7786.MCR-07-0267

have never been comprehensively studied. We therefore conducted a transcriptomic analysis of prospectively collected colorectal adenomas using a standardized oligonucleotide microarray covering the entire human genome. This study not only provided new information that is fundamental for future molecular characterization of these precancerous lesions but also allowed us to identify a putative marker of colorectal tumorigenesis.

Results

The focus of our study was the adenoma-adenocarcinoma pathway of colorectal carcinogenesis, which is closely linked to deregulation of the Wnt signaling pathway. To gain insight into the early steps of this process, we confined our investigation exclusively to sporadic, pedunculated colorectal adenomas (type 0-Ip of the Paris classification; ref. 6). Nonpolypoid and sessile polypoid lesions were not included because in some cases their transformation is believed to proceed along nonadenomatous pathways (7). Details on our case selection criteria are provided in Materials and Methods.

Thirty-two pedunculated adenomatous polyps, each with matched samples of normal mucosa, were prospectively collected from 28 patients (Table 1). The total number of synchronous and previously excised adenomas was <3 in 18 of 28 patients and 3 to 15 in the remaining 10. In this latter subgroup, the absence of *APC*- or *MYH*-associated multiple adenomatosis had been confirmed by genetic testing of lymphocyte DNA. Histologic analysis of one polyp (case NM) revealed superficial infiltration of the submucosa, but this case was not excluded because the region sampled for microarray analysis was clearly adenomatous. (As noted below, this finding was consistent with the results of hierarchical cluster analysis shown in Supplementary Fig. S1.).

Analysis of microarray data for the 32 adenoma/normal mucosa tissue pairs revealed that 31,033 of the probes were expressed in one or both of the tissue groups. The normal tissues were effectively segregated from the adenomas in four unsupervised analyses of the expression levels of these genes [hierarchical clustering, principal component analysis (PCA), correlation analysis, and correspondence analysis (CA); see Materials and Methods for details; Fig. 1]. In a separate

Table 1. Characteristics of the 28 Patients with Adenomatous Polyps Included in the Study

Patient	Age (y)	Sex	Colon segment involved	Maximum adenoma diameter (mm)	Microscopic appearance	Highest degree of dysplasia in the adenoma*	Degree of dysplasia at sampling site*	No. adenomas at study colonoscopy †	No. previously excised adenomas ‡	No. previous and/or synchronous hyperplastic polyps	Familiarity for colorectal cancer (relative, onset age)
GL§	49	M	D/S	10/10	T-V/T-V	H/L	H/L	9	4‖,¶	15**	Mother, 70
PR	74	F	S	20	T-V	H	H	2	—††	0	No
PC§	69	M	S/S	10/20	T/T-V	H/H	H/H	10	—	1	No
FP§	57	M	S/S	15/30	T/T-V	H/H	H/L	1	—	1	Mother, 69
CD§	71	M	T/R	15/10	T/T	H/L	H/L	2	7‖	2	No
MA	65	M	R	15	T-V	L	L	2	—	0	No
ME	63	M	R	15	T-V	L	L	9	—	1	Father, 79
RA	64	F	A	15	T	L	L	1	7‖	0	Sister, 68
PR	72	M	R	40	T-V	H	H	5	—	0	No
SD	56	M	A	15	T	H	H	1	1‖	0	Mother, 83; sister, 87
MP	38	M	S	15	T-V	L	L	2	—	0	Father, 79
MP	61	M	S	20	T	L	L	3	—	2	no
LG	41	M	R	20	T-V	H	L	5 (2 serrated)	—	0	Father, 60
LS	45	M	S	20	T	H	H	1	—	1	Father, 60
BG	58	M	D	15	T-V	L	L	2	—	1	No
PL	69	F	S	15	T-V	L	L	2	—	0	No
SMA	52	F	S	30	T	H	H	2	—	1	No
MR	58	F	D	20	T	L	L	2	—	0	No
GN	69	M	R	40	T	H	H	2	—	0	No
BA	69	M	S	30	T	L	L	6	—	0	No
PF	56	M	S	30	T	L	L	2	—	0	No
RC	55	F	A	30	T-V	L	L	12	3‖,¶	1	No
TMA	58	F	S	10	T	L	L	1	—	0	Mother, 85
NM	52	M	R	35	T-V	T1 ‡‡	H	1	—	0	No
MA	83	M	S	10	T-V	H	H	2	—	1	No
MM	50	M	S	30	T-V	H	H	2	—	0	Father's brother, 65
NF	79	M	A	20	T-V	L	L	2	—	0	Mother, 70
PN	67	F	S	15	T-V	L	L	1	—	1	No

Abbreviations: M, male; F, female; A, ascending colon; T, transversum; D, descending colon; S, sigmoid colon; R, rectum; T, tubular; T-V, tubulovillous; L, low-grade dysplasia; H, high-grade dysplasia.

*Low-grade versus high-grade dysplasia as defined by the WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6-9, 1999. IARC.

† This number includes the adenoma(s) subjected to microarray analysis.

‡ Total number of adenomas detected and excised during previous colonoscopies.

§Two adenomas from these patients were analyzed.

‖These cases were considered as recurrent adenomas for the CCA.

¶The index colonoscopy was done in a different center about 10 y before the study colonoscopy.

**Hyperplastic polyposis.

†† No previous colonoscopies.

‡‡ Superficial submucosal invasion (T1). The tissue collected for microarray came from the adenomatous portion of the polyp.

analysis, these two tissue groups were also unequivocally distinguished from a previously described set of 25 colon cancers (8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa (Supplementary Fig. S1).

Almost half of the expressed probes (15,059 of 31,033) displayed significant expression changes in adenomas. Those with fold changes ≥2 (1,190 probes up-regulated and 2,469 down-regulated in adenomas) were subjected to gene ontology analysis to identify the biological processes involved in the transition of normal mucosa to adenoma. The most significant results of this analysis are listed in Supplementary Table S1. The processes that were most markedly overrepresented among genes that were up-regulated in adenomas included mitosis, DNA replication, and spindle organization. Down-regulated genes were predominantly involved in host immune defense, inorganic anion transport, organ development, and inflammatory response, although a small group of genes involved in the latter process was up-regulated in adenomas (Supplementary Fig. S2).

We then analyzed the transcript levels of 319 genes believed to be components of the complex Wnt signaling pathway (Supplementary Table S2). Sixty-six of these genes (21%) were not expressed in either the normal or adenomatous tissue, and 34% were expressed similarly in both tissue groups. The remaining 144 genes displayed significantly altered expression in adenomas, and 78 of 144 displayed fold changes of ≥2.

A supervised extension of CA (9), canonical CA (CCA), was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. Four of the variables considered (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence; see Table 1) were clearly associated with distinct clusters of expression profiles (Fig. 2, variables in A and clusters for adenoma diameter in B; more details in the legend to this figure). The profile of adenomas measuring >20 mm could be easily distinguished from those of smaller (≤20 mm) adenomas. As shown by CCA and visualized on the corresponding CCA score plot (Fig. 2B), the centers of the three adenoma size clusters are distributed along the principal CCA axis (the vertical axis in Fig. 2B, the most important axis of separation of the expression profiles) in a definite order, with increasing diameters corresponding to progressively higher CCA scores. The variable large adenoma diameter was closely correlated with the vertical CCA axis (i.e., its vector "d>20mm" in Fig. 2A is almost parallel to this axis). It is interesting to note that the same correlation can be observed for the variable high-degree dysplasia (i.e., represented in Fig. 2A by vector "Hd"). This finding confirms the expected correlation between larger diameters and higher-degree dysplasia.

The CCA plot of the 11,709 modeled probes (loading plot, not shown) suggested that the distinction between the three size groups of adenomas is due to a complex network of relatively small changes in the expression of numerous genes (as opposed to marked changes involving a limited number of genes). Nevertheless, to maximize the use of the extensive data sets, we selected the 500 probes with the highest loading scores along the CCA axis 1 and isolated a set of genes whose expression changes displayed significant positive or negative correlation with adenoma size (Supplementary Table S3). Although their association with adenomas must be validated in a larger series, these are the expression changes most likely to play causal roles in the progression of these tumors.

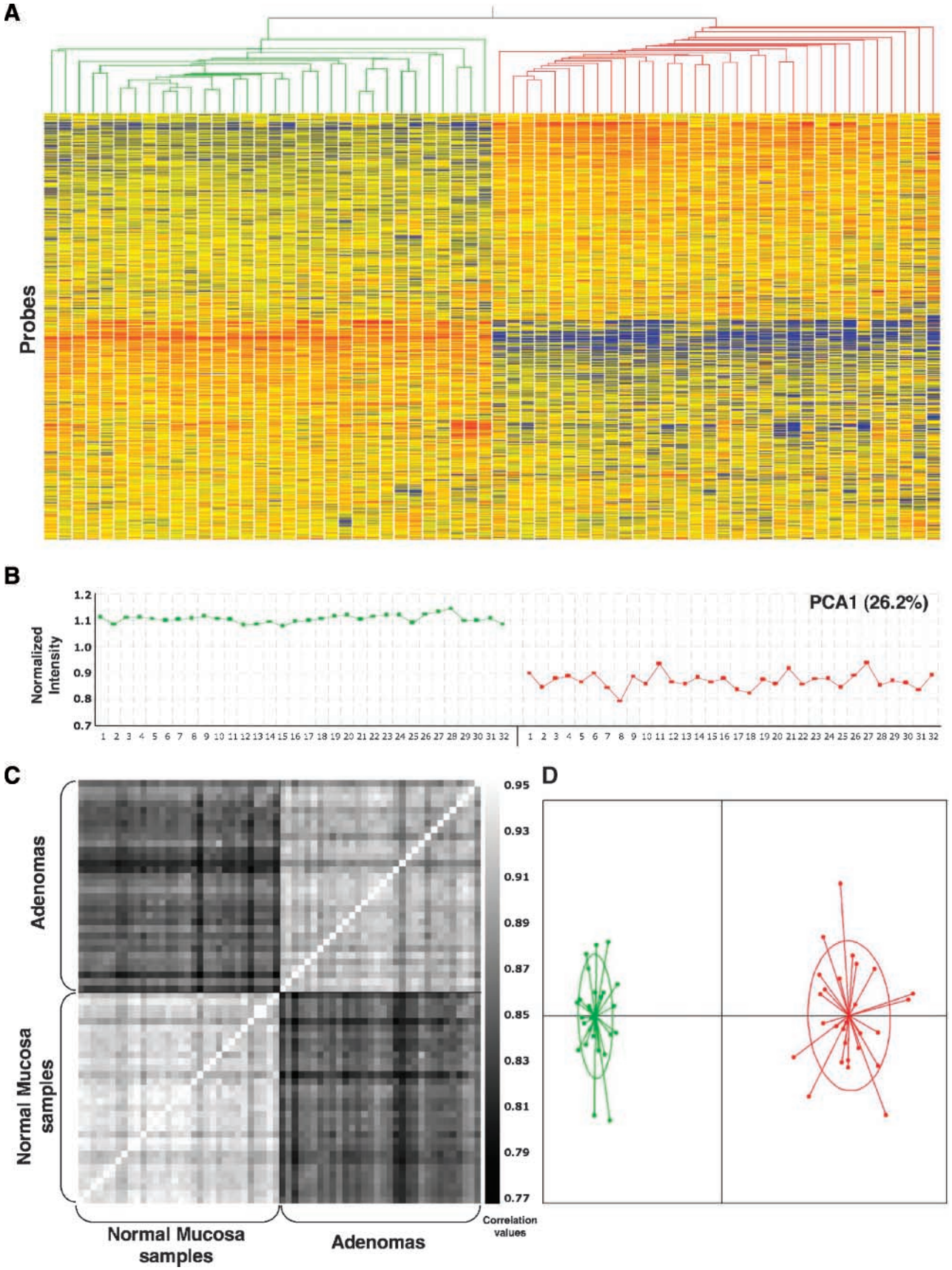
It should be mentioned that normal mucosa from the sigmoid colon had an expression profile that differed significantly from that of tissues from other colon segments (Fig. 2A). This finding will be explored in a future study conducted on a large series of normal mucosa samples from different colorectal segments.

The transcriptional profile of the 32 adenomas was thoroughly analyzed to identify genes likely to be involved in the development and evolution of these lesions. One of the first features that attracted our attention was the marked up-regulation of *KIAA1199* (Supplementary Table S4), a gene encoding a protein with unknown function. Its overexpression was striking in all colorectal adenomas we examined (average increases of 54.8-fold compared with normal mucosa) and in carcinomas (8). These findings were fully confirmed by real-time reverse transcription-PCR analysis of RNA extracted from samples used for the microarray study and from additional samples collected after the present study was completed (Supplementary Fig. S3).

In light of these findings, it was natural to wonder whether *KIAA1199* might be a novel positively regulated target of Wnt signaling, which is characteristically deregulated in colorectal tumors. Previous microarray studies indicated that genes coregulated at the transcriptional level under different conditions tend to be involved in the same processes and pathways, and the analysis of transcriptional coexpression has been used to predict the function of novel genes (10-12). Therefore, we conducted a search for known Wnt targets (listed in Supplementary Table S5) among the genes whose expression patterns in all the tissue samples significantly correlated with those of *KIAA1199*. (The procedure used in this analysis is summarized in Materials and Methods and Supplementary Fig. S4.) Forty-nine percent of the known Wnt targets that were overexpressed in our adenoma samples had expression patterns that were positively correlated with that of *KIAA1199* (Fig. 3A and B) as opposed to only 7.9% of the overexpressed genes that are not considered Wnt targets ($P < 0.0001$).

Evidence of the potential involvement of *KIAA1199* in the Wnt signaling pathway had also emerged from another study by our group (13). A combined analysis of microarray data of tissues and cell lines placed *KIAA1199* at the top of a list of genes [Supplementary Table S1 of ref. 13] that were up-regulated in colorectal adenomas and down-regulated in colon cancer cell lines that had undergone stable transfection with doxycycline-inducible forms of dominant-negative TCF1 or TCF4 to suppress Wnt signaling (14, 15). In the present study, *KIAA1199* was also found to be markedly down-regulated in LS174T colon cancer cells in which Wnt signaling had been blocked by the induction of β-catenin small interfering RNA or NH_2-terminal–deleted TCF4 (15, 16). The dramatic decrease in *KIAA1199* mRNA levels associated with this inhibition of the Wnt pathway was confirmed by Northern blotting (Fig. 3C).

In general, Wnt target genes are expressed predominantly in the proliferating compartment of normal intestinal crypts (lower portion), and their expression is appreciably increased in


A
Probes
B
Normalized Intensity
1.2
1.1
1.0
0.9
0.8
0.7
PCA1 (26.2%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
C
Adenomas
Normal Mucosa samples
0.95
0.93
0.91
0.89
0.87
0.85
0.83
0.81
0.79
0.77
Correlation values
Normal Mucosa samples
Adenomas
D

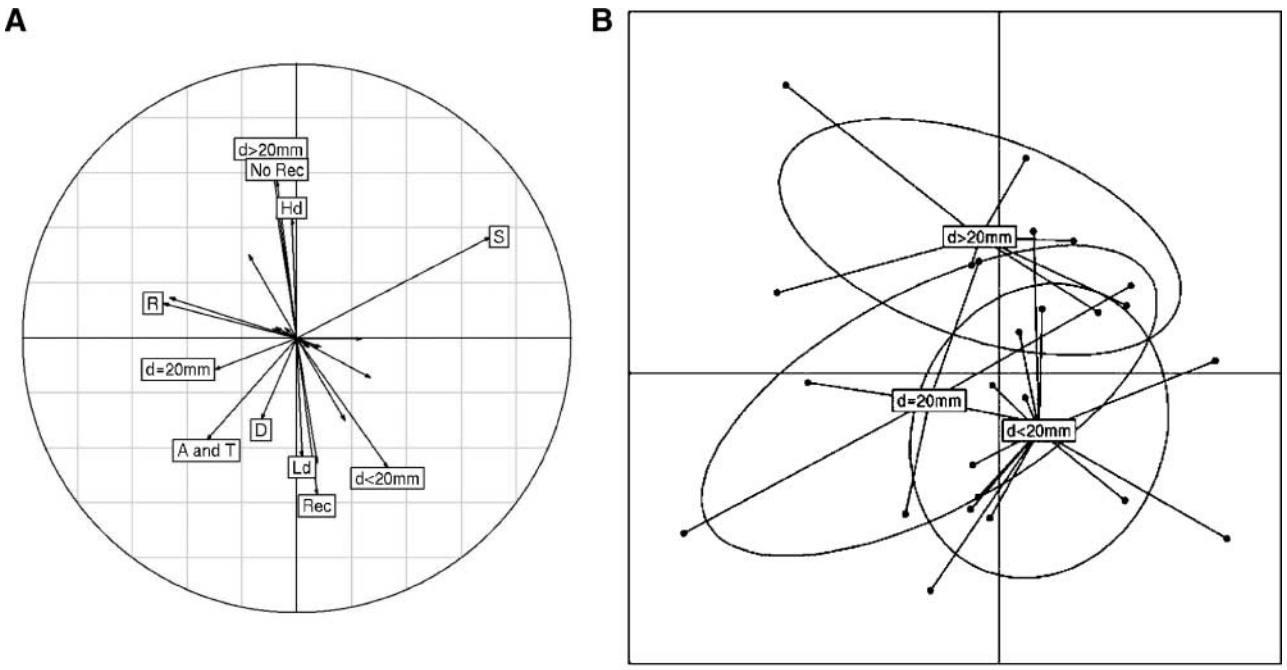


FIGURE 2. Clinical/pathologic variables that correlate with distinct gene expression profiles. The panels summarize the most important results of the CCA of mRNA intensity log-ratio values (adenoma: normal) of expressed genes. For clarity, CCA axis 1 has been drawn vertically in both panels. **A.** Correlation between specific clinical/pathologic variables (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence) and clusters of differential gene expression profiles (coded as log-ratio profiles), such as those shown in **B**. Each vector represents a specific value for a given variable (e.g., adenoma diameter of >20 mm and high-degree dysplasia) and points toward the center of the profile cluster correlated with the clinical/pathologic characteristic it represents. If the centers for each specific value are separated, the corresponding vectors point in distinct directions; otherwise, they are directed toward the same point. In the former case, the represented variable can be assumed to be significantly correlated with the profiles; in the latter case, there is no correlation. The length of the vector reflects the strength of the correlation: those approaching the circumference of the correlation circle, which represents a correlation value of 1, indicate stronger correlation than shorter vectors (correlation closer to 0). d, diameter; Hd, high-degree dysplasia; Ld, low-degree dysplasia; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; Rec, recurrent adenomas; no Rec, no recurrent adenomas. Unlabeled vectors are related to variables that were not clearly associated with any distinct cluster of expression profiles. Larger adenomas were predictably associated with high-degree dysplasia. In contrast, their association with nonrecurrence was unexpected and probably due to the fact that patients who had already undergone endoscopic polypectomy (i.e., those with recurrence) presented relatively recent-onset (consequently, smaller) polyps at the study colonoscopy. **B.** CCA score plot with samples grouped by adenoma diameter. Each of the three size-related groups is delimited by an ellipse with the center labeled. The ellipse representing the adenomas measuring >20 mm in diameter shows very little overlap with those of the other two groups (adenomas with diameters of 20 mm and those with diameters of <20 mm).

adenomatous glands (15). Our analysis of human tissues with preserved architecture indicated that these are also attributes of *KIAA1199*. In *in situ* hybridization studies, *KIAA1199* mRNA was detectable only in the lower portion of normal colonic epithelial crypts (Fig. 4A and B), and its expression levels were much higher in dysplastic glands (Fig. 4C). These patterns were confirmed at the protein level by immunohistochemistry done with an antibody raised in our laboratory (Fig. 4D-J). This analysis also revealed that the KIAA1199 is a cytoplasmic protein whose expression is most intense near the cell membrane, particularly on the luminal side of the dysplastic cell multilayer (Fig. 4F-J).

FIGURE 1. Unsupervised analyses of microarray data. **A.** Hierarchical clustering analysis. The 64 tissue samples represented on the *X* axis include 32 normal mucosal samples (*green branches*) and 32 adenomas (*red branches*). Each probe plotted on the *Y* axis is color coded to indicate the level of expression of the gene relative to its median expression level across the entire tissue sample set (*blue*, low; *red*, high). In the adenoma dendrogram, branches representing individual samples and small groups merge at higher levels than those of the normal mucosa dendrogram, reflecting lower-level correlation (i.e., higher variability among the adenoma specimens). **B.** PCA. Profile plot of the normalized first principal component (*PCA1*) across the 64 specimens (*green dots*, normal mucosa; *red dots*, adenomas). The two tissue groups differ significantly in terms of PCA1 ($P < 0.0001$), which accounted for 26% of the total variance. Note the higher variability of the PCA1 values in the adenoma group (higher fluctuation). **C.** Correlation analysis. Tile plot visualization of the pairwise correlations of the samples. Correlation values are indicated on the grayscale column (*white* > *black*: high > low). High correlation is observed among the samples within each group (*top right quadrant*, adenomas; *bottom left quadrant*, normal mucosa), although the adenomas displayed somewhat greater diversity (i.e., on the whole, the gray tones in the *top right quadrant* are darker than those in the *bottom left quadrant*). Top left and bottom right quadrants, normal and adenoma samples are poorly correlated. However, samples from the same patient generally showed higher correlation than that observed between normal and adenoma samples from different patients (bright pixels on the secondary diagonals in the *top left* and *bottom right quadrants*). This finding probably reflects the strong influence of several factors, including the individual genetic background and lifestyle and the fact that the normal and adenomatous tissues from a given patient were from the same colon segment. **D.** CA of mRNA log(intensity) values of expressed genes from 27 of the 32 tissue pairs (*green dots*, normal mucosa; *red dots*, adenoma). The other five pairs were excluded from this analysis because one of the two samples behaved as an outlier. Limiting our analysis to the more homogeneous pairs facilitated the comparison of the gene expression profiles for the two tissue groups and allowed more reliable identification of clinical/pathologic variables associated with profile scatter (see Fig. 2). The areas delimited by the ellipses represent 95% of the estimated binormal distribution of the sample scores on the first and second CA axes. The map of the sample scores on the first two axes shows that CA efficiently discriminates between normal and adenoma samples. Higher variability is evident in the adenoma group, where the samples are more widely dispersed.

Discussion

Adenomatous colorectal polyps are one of the most common human tumors and the most frequent precancerous lesions in the colorectum, but their transcriptome has been only partially analyzed, and the data are generally based on a limited number of cases (17-20). We attempted to fill this gap by doing a comprehensive whole-genome microarray analysis of a large, highly homogenous set of adenomas that was collected prospectively.

A comparison of the transcriptomes of adenomatous polyps and segment-matched samples of normal colorectal mucosa revealed evidence of broad-scale remodeling. As a starting point for future verification studies, we have drawn up a list of 478 genes that were significantly up-regulated (n = 153) or down-regulated (n = 325) in the adenomatous tissues (fold changes of ≥4; Supplementary Table S4). Space constraints preclude more than a cursory examination of this list, but we have highlighted in Table 2 certain aspects that we feel are particularly interesting in terms of their relevance to the process of adenoma formation. For instance, transcription regulation seems to be extensively modified. Twenty-nine molecules involved in this process were expressed in adenomas at levels

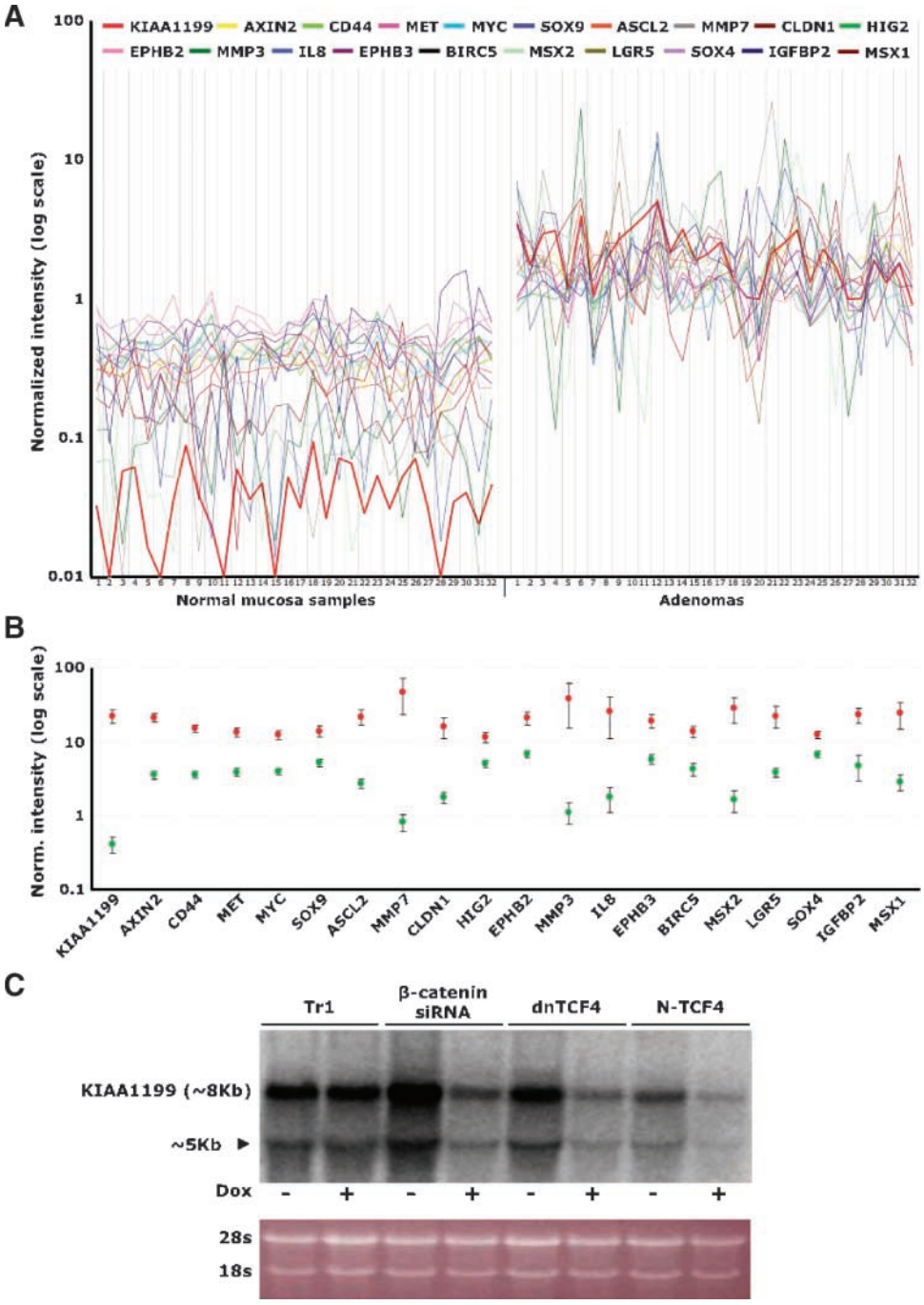


FIGURE 3. *KIAA1199* is a putative target of Wnt signaling. **A.** Degree of correlation between the expression of *KIAA1199* mRNA and that of 19 known Wnt signaling target genes identified with the procedure described in Materials and Methods, Results, and Supplementary Fig. S4. For each of the 20 genes, the graph shows the normalized intensity of expression level (plotted on the *Y axis*) in each of the 32 adenomas and corresponding samples of normal mucosa (*X axis*). **B.** Mean expression of each gene in normal mucosa (*green dots*) and adenomas (*red dots*). Bars, confidence interval. **C.** Northern blot showing reduced *KIAA1199* expression in LS174T cells following doxycycline-mediated induction of β-catenin small interfering RNA, dominant-negative TCF4 (*dnTCF4*), or NH_2-terminal–deleted TCF4 (*N-TCF4*). The ~8-kb band corresponds to full-length *KIAA1199* mRNA. The lower band (~5 kb) may represent an alternative form of this mRNA. Dox, cell transfectants grown in the presence or absence of doxycycline; Tr1, a parental clone (i.e., cells expressing the repressor protein modified by doxycycline but not transfected with β-catenin small interfering RNA, dominant-negative TCF4, or NH_2-terminal–deleted TCF4) used as a control of doxycycline exposure. Bottom, ethidium bromide–stained agarose gel as a loading control.

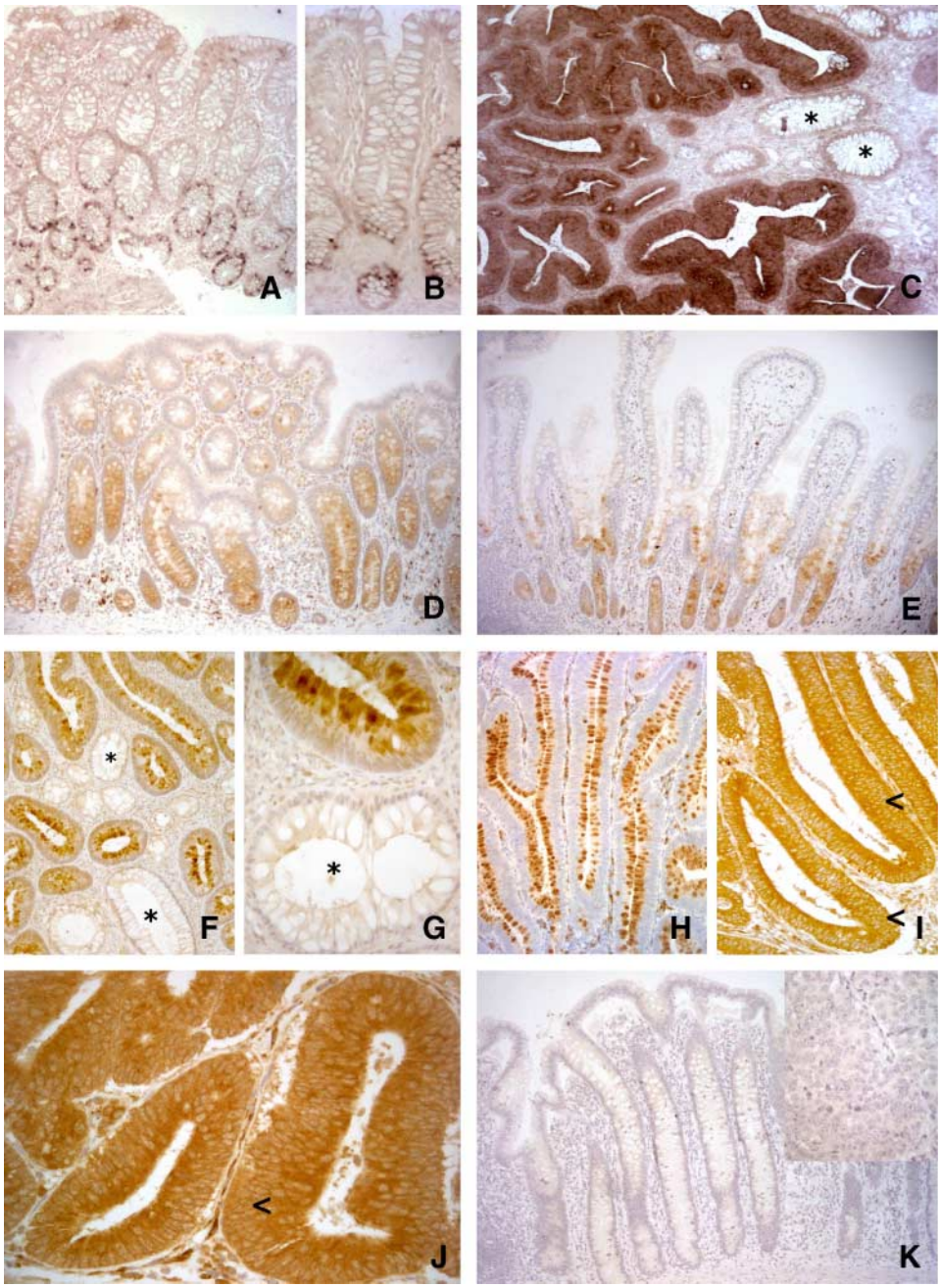


FIGURE 4. Expression of KIAA1199 mRNA and protein in normal intestinal mucosa and colorectal tumors. *In situ* hybridization studies (**A-C**) localized *KIAA1199* mRNA expression to the lower portion of normal epithelial crypts (**A** and **B**) and revealed that expression is markedly up-regulated in colorectal tumors (**C**). Asterisk, note the different levels of expression in tumor glands and normal crypts. **D.** KIAA1199 protein expression is also limited to the lower half of the normal colonic crypts, and a similar pattern is observed in the ileal mucosa (**E**), where the protein is expressed only in the crypts (not in the villi). In **F** and **G**, adenomatous crypts with low-grade dysplasia present increased expression of KIAA1199, particularly in the cytoplasm facing the crypt lumen, and in and around the mucin vacuoles of goblet cells (note the striking difference with goblet cells of normal crypts in both panels). The expression pattern changes dramatically during the transition from low-grade dysplasia with goblet cell differentiation (**H**) to high-grade dysplasia in which this differentiation is no longer apparent. **J.** In more advanced colon tumors, KIAA1199 overexpression is maintained. Note that, in **I** and **J**, the expression of KIAA1199 protein (like that of *KIAA1199* mRNA; **C**) is highest in the luminal portion of the dysplastic glands (*arrowheads*, multilayer of unstained nuclei occupying more than the basal half of the dysplastic epithelium). **K.** Normal mucosa, with the corresponding tumor in the inset. Negative control: KIAA1199 antibody preabsorbed with the peptide used to immunize rabbits.

>4 higher or lower than those observed in the normal mucosa, but there were also several smaller changes in this category (Supplementary Table S6) that might also have dramatic effects on gene expression. Several other alterations reported in Table 2 are noteworthy in terms of their potential effect on cell proliferation, differentiation, apoptosis, and cell adhesion: (*a*) up-regulation of four members of the REG (regenerating) family of genes (21, 22), which would lead to increased tissue mitogen expression; (*b*) up-regulation of *LCN2* (23) and down-regulation of *ZFHX1B/SIP-1* (24) in the absence of significant changes in the expression of the epithelial cadherin *CDH1* (E-cadherin), which would prevent or delay the epithelial-mesenchymal transition [changes were also noted in the expression of other cell adhesion genes of the cadherin and claudin families, including the striking overexpression of the placental cadherin gene *CDH3*, which is associated with early events in the transformation process (25, 26)]; (*c*) down-regulation of *ZFHX1B/SIP-1* and *Max dimerization protein 1* (*MXD1/MAD1*; decreased only 3.3-fold and therefore not listed in Table 2; refs. 27, 28) and overexpression of the *RTEL1* helicase, which should facilitate telomere elongation (29); (*d*) alterations that would diminish apoptosis [e.g., overexpression of the decoy receptor for Fas ligand, *TNFRSF6B*, which is reportedly coregulated with *RTEL1* on chromosome 20q13.3

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in Supplementary Table S4) Subdivided by Gene Ontology Category

Gene symbol	Gene name	Fold differences*	
		▲	▼
Regulation of transcription			
NLF1	Nuclear localized factor 1	33.1	
FOXQ1	Forkhead box Q1	24.4	
MSX2	Msh homeobox homologue 2	22.2	
ASCL2	Achaete-scute complex-like 2	17.3	
MSX1	Msh homeobox homologue 1	8.5	
IRX3	Iroquois homeobox protein 3	8.4	
GRHL3	Grainyhead-like 3	7.9	
TRIM29	Tripartite motif-containing 29	7.4	
ETV4	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	5.4	
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2	5.3	
TEAD4	TEA domain family member 4	5.2	
SP5	Sp5 transcription factor	5.2	
HES6	Hairy and enhancer of split 6	4.6	
TBX3	T-box 3	4.6	
NFE2L3	Nuclear factor (erythroid-derived 2)-like 3	4.3	
GRHL1	Grainyhead-like 1	4.2	
FEV	FEV (ETS oncogene family)		15.1
SPIB	Spi-B transcription factor		13.2
NEUROD1	Neurogenic differentiation 1		10.6
MEIS1	Meis1, myeloid ecotropic viral integration site 1		7.1
NR3C1	Nuclear receptor subfamily 3, group C, member 1		5.9
NR5A2	Nuclear receptor subfamily 5, group A, member 2		5.6
THRB	Thyroid hormone receptor, β		5.2
ZNF483	Zinc finger protein 483		5.1
ZFHX1B	Zinc finger homeobox 1b (SIP-1)		4.8
MEOX2	Mesenchyme homeobox 2		4.7
HOXD10	Homeobox D10		4.6
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene		4.5
SOX10	SRY (sex determining region Y)-box 10		4.2
Cell proliferation/differentiation/apoptosis			
REG1B	Regenerating islet-derived 1β	75.8	
REG3A	Regenerating islet-derived 3α	29.5	
TACSTD2	Tumor-associated calcium signal transducer 2	21.4	
IL-8	Interleukin-8	14.7	
SERPINB5	Serpin peptidase inhibitor, clade B, member 5 (Maspin)	13.8	
REG1A	Regenerating islet-derived 1α	8.2	
FAIM2	Fas apoptotic inhibitory molecule 2	7.5	
DUSP4	Dual specificity phosphatase 4	7.4	
REG4	Regenerating islet-derived family, member 4	6.8	
PHLDA1	Pleckstrin homology-like domain, family A, member 1	6.0	
LCN2	Lipocalin 2 (oncogene 24p3)	5.7	
RTEL1	Regulator of telomere elongation helicase 1	5.6	
TGFBI	Transforming growth factor, β induced	5.2	
IGFBP2	Insulin-like growth factor binding protein 2	4.8	
TDGF1	Teratocarcinoma-derived growth factor 1	4.7	
TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	4.5	
DMBT1	Deleted in malignant brain tumors 1	4.2	
TNFRSF10C	Tumor necrosis factor receptor superfamily, member 10c, decoy	4.1	
ANGPTL1	Angiopoietin-like 1 (Angioarrestin)		24.9
CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)		14.9
GPM6B	Glycoprotein M6B		11.5
ANK2	Ankyrin 2		9.8
UNC5C	Unc-5 homologue C		7.4
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)		6.1
CPNE8	Copine VIII		5.5
FAIM3	Fas apoptotic inhibitory molecule 3		5.4
IL6R	Interleukin-6 receptor		4.8
TUSC3	Tumor suppressor candidate 3		4.7
DUSP1	Dual specificity phosphatase 1		4.7
RERG	RAS-like, estrogen-regulated, growth inhibitor		4.6
NDN	Necdin		4.5
IGF1	Insulin-like growth factor 1 (somatomedin C)		4.0
Cell adhesion			
CDH3	Cadherin 3, type 1, P-cadherin	81.7	
CLDN2	Claudin 2	16.1	
CLDN1	Claudin 1	9.0	
DSG3	Desmoglein 3	7.3	

(*Continued on the following page*)

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in in Supplementary Table S4) Subdivided by Gene Ontology Category (Cont'd)

Gene symbol	Gene name	Fold differences*	
		▲	▼
DSG4	Desmoglein 4	5.9	
CLDN8	Claudin 8		25.8
CDH19	Cadherin 19, type 2		8.3
CEACAM7	Carcinoembryonic antigen-related cell adhesion molecule 7		8.3
CLDN23	Claudin 23		8.0
NRXN1	Neurexin 1		7.1
PCDH19	Protocadherin 19		6.8
NLGN4X	Neuroligin 4, X-linked		6.0
TNXB	Tenascin XB		5.6
MUCDHL	Mucin and cadherin-like		5.1
PCDH9	Protocadherin 9		4.9
L1CAM	L1 cell adhesion molecule		4.2

*Overexpressed (▲) or underexpressed (▼) in adenomas (versus normal mucosa samples).

(30-32); decreased expression of the netrin-1 receptor, *UNC5C* (33); and expression changes involving three Fas apoptosis inhibitory molecules (*FAIM*), including *FAIM1*, which was increased 2.3-fold and is thus not listed in Table 2]; and (*e*) marked down-regulation of several genes that would result in reduced tumor suppression activity [e.g., those encoding the antiangiogenic factor ANGPTL1 (34), the cyclin-dependent kinase inhibitor CDKN2B/p15, and the prostaglandin catabolism enzyme HPGD (35)].

It is also important to recall the size-related differences noted in the adenoma gene expression profiles (Fig. 2; Supplementary Table S3). When validated in a larger series of tumors, these differences should provide important clues to the molecular basis of the well-known link between the dimensions and malignant potential of colorectal adenomas (1).

Our study also furnishes a complete picture of expression changes involving gene components of the Wnt pathway across the transition from normal to adenomatous epithelium (Supplementary Table S2) as well as evidence for the existence of a novel Wnt target: *KIAA1199*. This gene, which encodes a protein of unknown function, was strikingly overexpressed in all the adenomas included in this study and in 25 adenocarcinomas of the colon described in a previous report (8). Even more intriguingly, its expression was significantly correlated with that of several genes that are well-established targets of Wnt signaling. Our hypothesis that *KIAA1199* is up-regulated by the TCF(s)/β-catenin transcription complex was considerably strengthened by the marked decreases in *KIAA1199* expression observed in cultured colorectal cancer cells when the Wnt pathway was inhibited by overexpression of dominant-negative TCF4 proteins or by β-catenin knockdown. It is not yet clear whether this is a direct effect, but this possibility is supported by the results of a recent genome-wide TCF4 ChIP-on-chip analysis, which indicates that the *KIAA1199* locus is surrounded by four TCF4-bound regions.[10] These findings are consistent with the probable role of this gene as a direct target of TCF4/β-catenin signaling in the intestine and in colorectal tumors.

[10] Hatzis et al., unpublished data.

Other features of *KIAA1199* expression are also compatible with its putative role as a Wnt target gene. *KIAA1199* mRNA and protein are both confined to the proliferative compartment of normal intestinal crypts, where Wnt signaling is normally active, and they are highly overexpressed in colorectal adenomas and carcinomas, where this pathway is almost always aberrantly activated.

In normal and tumor tissues, KIAA1199 is expressed in the cytoplasm of epithelial cells. In glands with low-degree dysplasia, higher concentrations are observed in the mucin vacuoles of goblet cells, but cytoplasmic expression of the protein in tumor cells remains elevated even after goblet cell differentiation has been lost (Fig. 4). These features, together with the localization of KIAA1199 in the luminal portion of the cytoplasm, are suggestive of a secreted and/or membrane protein. This conclusion is consistent with our *in silico* analysis of KIAA1199 (see Supplementary Data and Supplementary Fig. S5), which strongly predicts the presence of a signal peptide at its NH_2-terminal end. In addition, the central region of KIAA1199 contains a TMEM2 homology domain, which is present in several eukaryotic proteins, including TMEM2, polyductin (PKHD1), and fibrocystin L (PKHD1L1; Fig. 5), all large receptor proteins characterized by an NH_2-terminal signal peptide or a single transmembrane helix and a short cytoplasmic tail (36).

A study based on yeast two-hybrid screens suggested that KIAA1199 may interact with plexin A2 (KIAA0463; ref. 37). The transmembrane plexins interact with transmembrane semaphorins on nearby cells, providing "stop" and "go" signals that are crucial for cell motility and invasive growth (38, 39). KIAA1199/plexin A2 interaction could thus play important roles in colorectal tumorigenesis not only in the invasive stages but also earlier during the formation of abnormal glands in benign adenomas.

A recent report linked high levels of *KIAA1199* mRNA with cell mortality in human fibroblasts and in a renal cell carcinoma cell line (40). In that study, however, there was no significant increase in *KIAA1199* expression during replicative aging of mortal cells, and this finding contrasts with the documented behavior of other genes involved in cell aging (41). Furthermore,

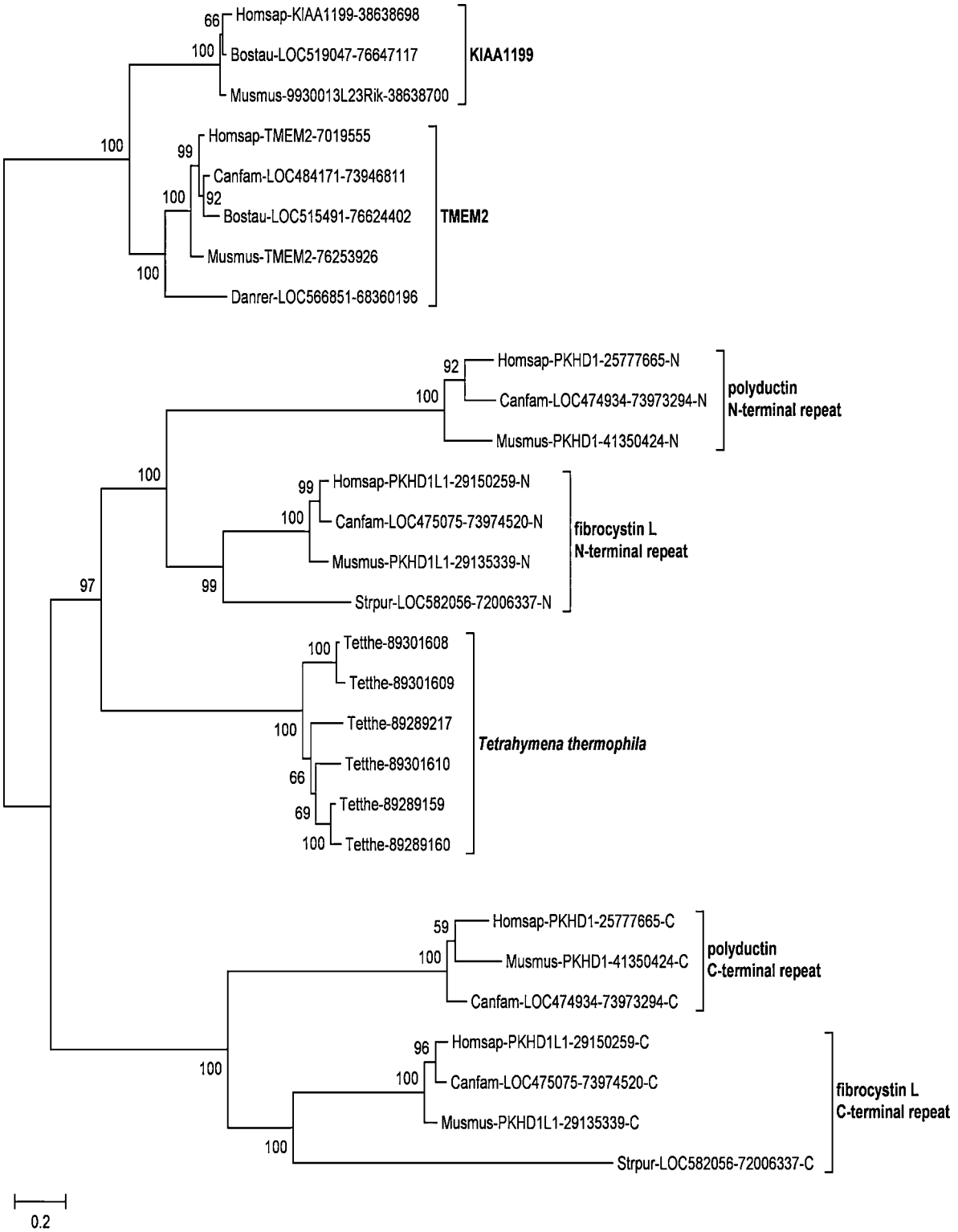


FIGURE 5. Phylogenetic tree of the proteins containing the TMEM2 homology domain found in the central region of KIAA1199. The tree was generated with MEGA3 (52) from the multiple sequence alignment shown in Supplementary Fig. S5. It was calculated with the minimum evolution algorithm and the JTT matrix. Positions with gaps were removed for calculation of pairwise distances. Node robustness was assessed using the bootstrap method with 100 resamplings. (Bootstrap values are shown at the nodes.) Two branches emerged, one comprising KIAA1199 and TMEM2 and the other with polyductin, fibrocystin L, and several other THD-containing proteins found in the ciliate *Tetrahymena thermophila*, which were apparently generated in a series of *Tetrahymena*-specific gene duplications. The NH_2-terminal repeats of polyductin and fibrocystin L clustered together, as did the COOH-terminal repeats, suggesting that the intragenic duplication of the TH domain in the ancestor of polyductin and fibrocystin L occurred before the divergence of chordates and echinoderms (more details in Supplementary Data).

the authors reported wide variation in *KIAA1199* mRNA expression in breast cancer cell lines, and this finding raises the possibility that expression of this gene *in vivo* and in cell lines may differ.

We believe that our microarray data will serve as a springboard and reference point for other studies on the molecular basis of colorectal transformation along the adenoma-carcinoma pathway (and subsequently for the study of alternative pathways). Some of the transcriptional changes reported in this study might one day be used as molecular indices of the susceptibility of adenomas to malignant transformation, information that would be helpful in planning appropriate follow-up of the lesions. As for KIAA1199, its invariably high expression in the colorectal tumors we studied raises interesting possibilities for the development of a new molecular marker for the detection of these neoplasms. For example, because KIAA1199 expression in the normal mucosa is limited to cells in the lower portion of the crypts, which are not yet programmed to be shed into the intestinal lumen, the presence of KIAA1199 peptides in fecal water might prove to be a specific marker of adenomatous lesions. In addition, although due consideration must be given to its probable physiologic role(s) in intestinal crypts and possibly in several other human tissues (40, 42, 43), KIAA1199 may be a potential target of antibody-based therapies.

Materials and Methods

Tumor Samples

Pedunculated colorectal polyps and normal mucosa were obtained during colonoscopies carried out in the Gastroenterology Unit of the Belcolle City Hospital (Viterbo, Italy). The tissues were collected prospectively with informed patient consent and the approval of the local Human Research Ethics Committee. Patients with documented familial polyposis, with >15 adenomatous polyps (total: synchronous + previously excised; ref. 44), or currently treated with nonsteroidal anti-inflammatory drugs (including aspirin) were excluded from the study.

For each polyp, three biopsies of normal mucosa were collected from the same colon segment (≥2 cm from the site of the polyp). Immediately after removal, a small sample of epithelial tissue (5-15 mg) was cut from the tip of each polyp, leaving the underlying muscularis mucosae intact. We excluded polyps <1 cm to ensure that the sampling procedure would not interfere with the histologic diagnosis. All polyp samples were collected by a single operator (M.d.P.) using the same procedure to minimize artifacts due to sampling differences. The approach used allowed us to obtain specimens with a high percentage of epithelial cells without resorting to microdissection, which can diminish the quantity and quality of the extracted RNA.

The polyp sample and the three normal mucosal biopsies were immersed in RNAlater (Ambion) for subsequent microarray analysis, and the remainder of the polyp was submitted for pathologic analysis. The cut surface at the tip was labeled with India ink so that the sampled area could be easily identified during routine histologic examination. The tissue was then fixed in buffered formalin and embedded in paraffin. DNA extracted from sections of this specimen was also used to rule out microsatellite instability (reflecting defective DNA mismatch repair) at the *BAT26* locus, as previously described (45). All of the polyps included in the study met the following criteria: type 0-Ip (6), maximum diameter of 1 to 4 cm, absence of surface ulceration, histologic diagnosis of adenoma, and absence of microsatellite instability at *BAT26*.

In some analyses, we also included transcriptomic data from a previously described set of 25 colon cancers (mismatch repair proficient and deficient; ref. 8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa.

Microarray Analysis, Real-time Reverse Transcription-PCR, and Northern Blotting

Total RNA was extracted (RNeasy Mini kit, Qiagen) from homogenized tissue samples (5-15 mg), and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies). Complementary RNA (15 μg/sample), synthesized and labeled as previously described (8, 46), was hybridized with the Affymetrix U133 Plus 2.0 array, which contains *in situ* synthesized oligonucleotides representing the entire human genome (54,675 probes).

Raw gene expression data generated by GeneChip Operating Software (Affymetrix) were imported into the GeneSpring software program (Agilent Technologies) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples). Analysis was done using the log expression values with GeneSpring's cross-gene error model turned on. Probes were excluded from analysis unless they were listed as "present or marginal calls" and/or had expression values ≥100 in ≥50% (≥16 of 32) of the samples in at least one of the tissue groups (adenomas and normal mucosa).

Expression data were subjected to four different unsupervised analyses: (*a*) hierarchical clustering using the Pearson correlation coefficient as a similarity measure and the average linkage algorithm for branch merging; (*b*) PCA, which reduces the dimensionality (number of variables) of a data set while retaining most of its variance (8); (*c*) correlation analysis, which involved computation of Pearson correlation coefficients for all possible sample pairs and visualization of correlation values as tile plots; and (*d*) CA, another dimension-reducing method (47), which was used to identify samples associated with particular gene expression levels. In typical CA, a matrix of *n* gene expression levels from *p* samples is treated as a two-way contingency table (genes by samples or vice versa) with *n* and *p* specifications for the "factors" gene and sample, respectively. Each intensity value thus reflects the abundance of a given transcript in a given sample. Like PCA, CA identifies independent "factorial components" that account for variance within a multidimensional gene data set, but in this case, the components are identified and ranked according to the correlation between gene and sample scores. A supervised or constrained extension of CA (9), CCA, was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. CA and CCA, as well as the corresponding plots, were computed using R software and the ade4 and made4 packages furnished by Bioconductor.[11]

[11] http://www.bioconductor.org

The Mann-Whitney test was used to select genes differentially expressed in normal mucosa and adenomas; Benjamini-Hochberg multiple testing correction was applied with a false discovery rate of 0.01. The genes in this set that were differentially expressed with fold differences of $\geq$2.0 were then analyzed with ErmineJ software (48) to identify any biological processes from the Gene Ontology database (49) that were overrepresented.

Pearson correlation was used to identify correlation between *KIAA1199* expression and the expression of other genes in the entire set of tissue samples. Fisher's exact test was used to identify possible overrepresentation of known Wnt targets among genes whose expression was closely correlated with that of *KIAA1199* (correlation values $\geq$0.8).

Reverse transcription-PCR and Northern blotting were done as previously described (46, 50) to verify the expression level of *KIAA1199* in tissue samples and in LS174T colon cancer cells in which inducible inhibition of the Wnt pathway had been achieved with previously described methods (14-16).

In situ *Hybridization*

Digoxigenin-labeled *KIAA1199* antisense riboprobes were synthesized from a PCR product amplified from human colon cDNA with *KIAA1199*-specific primers (sense: 5′-cacatcgggg-gaggagataga-3′; antisense, containing a T7 RNA polymerase-binding site: 5′-taatacgactcactatagggttccagacttgaca-3′). This product was transcribed *in vitro* using the DIG RNA labeling kit and T7 RNA polymerase (Roche Diagnostics). *In situ* hybridizations were done on paraffin-embedded sections of human colon fixed with 4% buffered formalin as described elsewhere (51).

Immunohistochemistry

Our *in silico* analysis of KIAA1199 (see Supplementary Data) indicated that residues 202 to 217 (IHSDRFDTYRSKKESE) form a loop between a conserved β-strand and the following helix of the NH_2-terminal GG domain. This charged, surface-exposed peptide was used to raise a rabbit polyclonal antibody, which was purified by affinity chromatography on Thiopropyl Sepharose 6B (Amersham) derivatized with the antigenic peptide. A 1:1,000 dilution of this antibody was used, as previously described (45), to evaluate KIAA1199 expression in formalin-fixed, paraffin-embedded sections of adenoma and normal mucosal tissues.

Acknowledgments

We thank the patients who participated in this study; R. Haider, E. Pani, T. Lehmann, A. Patrignani, and U. Wagner for technical assistance; A. Mueller, E. Veljkovic, and F. Bannwart for critical reading of the manuscript; and M. Kent for editorial assistance.

References

1. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. Gastroenterology 2006;130:1872–85.

2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology 1987;93:1009–13.

3. Pinto D, Clevers H. Wnt, stem cells and cancer in the intestine. Biol Cell 2005; 97:185–96.

4. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a β-catenin-Tcf complex in APC−/− colon carcinoma. Science 1997;275: 1784–7.

5. Morin PJ, Sparks AB, Korinek V, et al. Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC. Science 1997;275:1787–90.

6. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570–8.

7. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113–30.

8. di Pietro M, Sabates Bellver J, Menigatti M, et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–59.

9. Ter Braak C. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 1986;67:1167–79.

10. Miki R, Kadota K, Bono H, et al. Delineating developmental and metabolic pathways *in vivo* by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc Natl Acad Sci U S A 2001;98:2199–204.

11. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ. Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. Nat Genet 2002;31:255–65.

12. Zhang W, Morris QD, Chang R, et al. The functional landscape of mouse gene expression. J Biol 2004;3:21.

13. Van der Flier LG, Sabates-Bellver J, Oving I, et al. The intestinal Wnt/TCF signature. Gastroenterology 2007;132:628–32.

14. van de Wetering M, Barker N, Harkes IC, et al. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61: 278–84.

15. van de Wetering M, Sancho E, Verweij C, et al. The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–50.

16. van de Wetering M, Oving I, Muncan V, et al. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO Rep 2003;4:609–15.

17. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124–30.

18. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21: 4120–8.

19. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K. Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. Br J Cancer 2005;92:1193–200.

20. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. Gut 2003;52:1148–53.

21. Nata K, Liu Y, Xu L, et al. Molecular cloning, expression and chromosomal localization of a novel human REG family gene, REG III. Gene 2004;340:161–70.

22. Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK. Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. Gastroenterology 2006;130:137–49.

23. Hanai J, Mammoto T, Seth P, et al. Lipocalin 2 diminishes invasiveness and metastasis of Ras-transformed cells. J Biol Chem 2005;280:13641–7.

24. Comijn J, Berx G, Vermassen P, et al. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell 2001;7:1267–78.

25. Jankowski JA, Bedford FK, Boulton RA, et al. Alterations in classical cadherins associated with progression in ulcerative and Crohn's colitis. Lab Invest 1998;78:1155–67.

26. Hardy RG, Tselepis C, Hoyland J, et al. Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon. Gut 2002; 50:513–9.

27. Lin SY, Elledge SJ. Multiple tumor suppressor pathways negatively regulate telomerase. Cell 2003;113:881–9.

28. Mariadason JM, Nicholas C, L'Italien KE, et al. Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. Gastroenterology 2005;128:1081–8.

29. Ding H, Schertzer M, Wu X, et al. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. Cell 2004;117:873–86.

30. Pitti RM, Marsters SA, Lawrence DA, et al. Genomic amplification of a

decoy receptor for Fas ligand in lung and colon cancer. Nature 1998;396: 699–703.

31. Bai C, Connolly B, Metzker ML, et al. Overexpression of M68/DcR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster. Proc Natl Acad Sci U S A 2000;97:1230–5.

32. Postma C, Hermsen MA, Coffa J, et al. Chromosomal instability in flat adenomas and carcinomas of the colon. J Pathol 2005;205:514–21.

33. Thiebault K, Mazelin L, Pays L, et al. The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. Proc Natl Acad Sci U S A 2003;100:4173–8.

34. Dhanabal M, Jeffers M, LaRochelle WJ, Lichenstein HS. Angioarrestin: a unique angiopoietin-related protein with anti-angiogenic properties. Biochem Biophys Res Commun 2005;333:308–15.

35. Backlund MG, Mann JR, Holla VR, et al. 15-Hydroxyprostaglandin dehydrogenase is down-regulated in colorectal cancer. J Biol Chem 2005;280: 3217–23.

36. Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. Hum Mol Genet 2003;12:685–98.

37. Nakayama M, Kikuno R, Ohara O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. Genome Res 2002;12:1773–84.

38. Tamagnone L, Artigiani S, Chen H, et al. Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. Cell 1999;99:71–80.

39. Comoglio PM, Tamagnone L, Giordano S. Invasive growth: a two-way street for semaphorin signalling. Nat Cell Biol 2004;6:1155–7.

40. Michishita E, Garces G, Barrett JC, Horikawa I. Upregulation of the KIAA1199 gene is associated with cellular mortality. Cancer Lett 2006;239: 71–7.

41. Horikawa I, Parker ES, Solomon GG, Barrett JC. Upregulation of the gene encoding a cytoplasmic dynein intermediate chain in senescent human cells. J Cell Biochem 2001;82:415–21.

42. Abe S, Katagiri T, Saito-Hisaminato A, et al. Identification of CRYM as a candidate responsible for nonsyndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues. Am J Hum Genet 2003;72: 73–82.

43. Abe S, Usami S, Nakamura Y. Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. J Hum Genet 2003;48:564–70.

44. Marra G, Jiricny J. Multiple colorectal adenomas—is their number up? N Engl J Med 2003;348:845–7.

45. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160–71.

46. di Pietro M, Marra G, Cejka P, et al. Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine. Cancer Res 2003;63:8158–66.

47. Benzécri J-P. Correspondence analysis handbook. New York: Marcel Dekker, Inc.; 1992.

48. Lee HK, Braynen W, Keshav K, Pavlidis P. ErmineJ: tool for functional analysis of gene expression data sets. BMC Bioinformatics 2005;6:269.

49. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

50. Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H. The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. EMBO J 2001;20:4935–43.

51. Maake C, Auf der Maur F, Jovanovic K, Reinecke M, Hauri D, John H. Occurrence and localization of uroguanylin in the aging human prostate. Histochem Cell Biol 2003;119:69–76.

52. Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004;5: 150–63.

Chapter 4

Genome-Wide Pattern of TCF7L2/TCF4 Chromatin Occupancy in Colorectal Cancer Cells

Reprinted from: Mol Cell Biol. 2008 Apr;28(8):2732-44

Genome-Wide Pattern of TCF7L2/TCF4 Chromatin Occupancy in Colorectal Cancer Cells▽†

Pantelis Hatzis,[1] Laurens G. van der Flier,[1] Marc A. van Driel,[2] Victor Guryev,[1] Fiona Nielsen,[2] Sergei Denissov,[2] Isaäc J. Nijman,[1] Jan Koster,[3] Evan E. Santo,[3] Willem Welboren,[2] Rogier Versteeg,[3] Edwin Cuppen,[1] Marc van de Wetering,[1] Hans Clevers,[1]‡* and Hendrik G. Stunnenberg[2]‡*

Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands[1]; Nijmegen Center for Molecular Life Sciences, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands[2]; and Department of Human Genetics, M1-134, Academic Medical Center, University of Amsterdam, P.O. Box 22700, 1100 DE Amsterdam, The Netherlands[3]

Received 7 December 2007/Returned for modification 17 January 2008/Accepted 4 February 2008

Wnt signaling activates gene expression through the induced formation of complexes between DNA-binding T-cell factors (TCFs) and the transcriptional coactivator β-catenin. In colorectal cancer, activating Wnt pathway mutations transform epithelial cells through the inappropriate activation of a TCF7L2/TCF4 target gene program. Through a DNA array-based genome-wide analysis of TCF4 chromatin occupancy, we have identified 6,868 high-confidence TCF4-binding sites in the LS174T colorectal cancer cell line. Most TCF4-binding sites are located at large distances from transcription start sites, while target genes are frequently "decorated" by multiple binding sites. Motif discovery algorithms define the in vivo-occupied TCF4-binding site as evolutionarily conserved A-C/G-A/T-T-C-A-A-A-G motifs. The TCF4-binding regions significantly correlate with Wnt-responsive gene expression profiles derived from primary human adenomas and often behave as β-catenin/TCF4-dependent enhancers in transient reporter assays.

Physiological Wnt signaling is required for the maintenance of the crypt progenitor phenotype and controls the proliferation/differentiation switch in the adult, self-renewing intestinal epithelium (33). A constitutively active Tcf/β-catenin transcription complex, resulting from mutations in adenomatous polyposis coli (APC), Axin, or β-catenin, is the primary transforming factor in colorectal cancer (CRC) (25, 26, 32); aberrant Tcf/β-catenin activity results in a transcriptional profile in CRC cells similar to that which is physiologically driven by Tcf/β-catenin in the crypt stem/progenitor cells of the intestine (49). Through candidate gene approaches and microarray technology, a large number of genes have been uncovered whose expression levels are altered upon abrogation or activation of the Wnt pathway (for references, see http://www.stanford.edu/~rnusse/pathways/targets.html). It remains unclear whether the affected genes are direct or indirect targets of the Tcf/β-catenin transcription factor complex. *cis*-regulatory elements directly bound by Tcf have been identified for only a few candidate genes. Such studies have been mostly limited to regulatory regions close to the transcription start site (TSS) of candidate genes (e.g., see reference 17). A comprehensive identification of regulatory elements is essential for a more complete understanding of the transcriptional repertoire driven by the Wnt pathway and the elucidation of the molecular mechanisms by which Tcf and β-catenin control the transcription of their target genes.

A recent approach taken to achieve such goals is chromatin immunoprecipitation (ChIP)-coupled DNA microarray analysis (ChIP-on-chip), which couples the immunoprecipitation of chromatin-bound transcription factors with the identification of the bound DNA sequences through hybridization on DNA microarrays (35). This approach has been used to generate, among others, a comprehensive map of active, preinitiation complex-bound promoters in human fibroblast cells (24). Microarrays covering the nonrepetitive sequence of chromosomes 21 and 22 have allowed the study of histone H3 methylation and acetylation patterns in human hepatoma cells (5) and estrogen receptor binding sites in breast cancer cells (8). The latter study revealed selective binding of estrogen receptor (ER) to a limited number of sites, most of which were distant from the TSSs of ER-regulated genes (8). Similar conclusions were put forth by work examining the in vivo binding of transcription factors Sp1, c-Myc, and p53 along chromosomes 21 and 22: most binding sites identified do not correspond to the proximal promoters of protein-coding genes but rather lie within or immediately 3′ to well-characterized genes or are significantly correlated with noncoding RNAs (10). Collectively these studies point to the necessity of interrogating entire genomes for the comprehensive determination of in vivo-occupied binding sites (9, 23, 52, 54).

In the present work, we used a combination of ChIP and location analysis with genome-wide tiling arrays to generate a genome-wide binding profile of TCF4, the T-cell factor (TCF)

* Corresponding author. Mailing address for Hans Clevers: Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. Phone: 31-302-12-1831. Fax: 31-302-12-1801. E-mail: h.clevers@niob.knaw.nl. Mailing address for Hendrik G. Stunnenberg: Nijmegen Center for Molecular Life Sciences, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands. Phone: 31-24-3610524. Fax: 31-24-3610520. E-mail: h.stunnenberg@ncmls.ru.nl.

‡ These authors contributed equally.

† Supplemental material for this article may be found at http://mcb.asm.org/.

▽ Published ahead of print on 11 February 2008.

family member most prominently expressed in the mammalian intestine (1, 26).

MATERIALS AND METHODS

ChIP. LS174T cells were cross-linked with 1% formaldehyde for 20 min at room temperature. The reaction was quenched with glycine at a final concentration of 0.125 M. The cells were successively washed with phosphate-buffered saline, buffer B (0.25% Triton-X 100, 10 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]) and buffer C (0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]) at 4°C for 10 min each. The cells were then resuspended in ChIP incubation buffer (0.3% sodium dodecyl sulfate [SDS], 1% Triton-X 100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]) and sheared using a Bioruptor sonicator (Cosmo Bio Co., Ltd.) with six pulses of 30 s each at the maximum setting. The sonicated chromatin was centrifuged for 15 min and incubated for 12 h at 4°C with either a polyclonal anti-TCF4 antibody (sc-8631; Santa Cruz Biotechnology, Inc.) or a monoclonal anti-TCF4 antibody (1) (05-511; Upstate) at 1 μg of antibody per 10^6 cells with protein G beads (Upstate). The beads were successively washed 2 times with buffer 1 (0.1% SDS, 0.1% deoxycholate, 1% Triton-X 100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]), one time with buffer 2 (0.1% SDS, 0.1% sodium deoxycholate, 1% Triton-X 100, 0.5 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]), one time with buffer 3 (0.25 M LiCl, 0.5% sodium deoxycholate, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]), and two times with buffer 4 (1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES [pH 7.6]) for 5 min each at 4°C. The precipitated chromatin was eluted by incubation of the beads with elution buffer (1% SDS, 0.1 M $NaHCO_3$) at room temperature for 20 min, de-cross-linked by incubation at 65°C for 5 h in the presence of 200 mM NaCl, extracted with phenol-chloroform, and precipitated.

For sequential ChIP, the eluted chromatin was diluted with ChIP incubation buffer without SDS to the incubation conditions of the first ChIP. Half the amount of antibody was added to the second ChIP and processed as for the first.

Ligation-mediated PCR amplification, labeling, and hybridization. The ChIP material was amplified for labeling as described previously (35). Labeling of the material, hybridization, and scanning of the arrays were performed by Nimblegen, Inc.

Quantitative PCR (qPCR). ChIP experiments were analyzed with quantitative PCR in an iCycler iQ real-time PCR detection system (Bio-Rad), using iQ Sybr green supermix (Bio-Rad). Specific primers were designed using Beacon Designer software (Premier Biosoft International) and verified for specificity by in silico PCR (http://genome.cse.ucsc.edu/cgi-bin/hgPcr). ChIP values were normalized as a percentage of input. The specificity of ChIP values was expressed as the change from respective values for control regions (i.e., exon 2 of the nonexpressed myoglobin gene). Based on TCF4 occupancy values over a number of such negative control regions, we defined as positive those regions whose change in occupancy over the control region was greater than threefold.

Reporter assays. Genomic fragments encompassing typically about 1 kb of genomic sequence encompassing a TCF4 peak were amplified by PCR from human genomic DNA and cloned in front of the firefly luciferase gene in pGL3b or pGL4.10, in the case of TSS-proximal regions, or in front of a minimal fragment encompassing the TATA box of the adenovirus major late promoter cloned in front of the firefly luciferase gene in pGL3b or a minimal TATA box cloned in front of the firefly luciferase gene in pGL4.10, in the case of non-TSS-proximal regions. For the control experiment, human genomic DNA was digested with KpnI and 15 fragments of approximately 1 kb cloned in front of the firefly luciferase gene in pGL3b and 15 fragments cloned in front of a minimal fragment encompassing the TATA box of the adenovirus major late promoter cloned in front of the firefly luciferase gene in pGL3b were used in the reporter assays. The reporters were transfected with Fugene 6 (Roche Diagnostics) in LS174T or LS174T/ΔNTCF4 cells (the latter inducibly overexpress ΔNTCF4 upon doxycycline treatment) with *Renilla* luciferase as a transfection control and appropriate expression vectors, and their activity was measured using a dual-luciferase reporter assay system (Promega).

Array design. The genome-wide hybridization was performed on a NimbleGen Systems, Inc., set of 36 arrays containing a total of 13,787,634 oligonucleotides of 50 bp covering the repeat-masked portion of the human genome for chromosomes 1 to 22p at 100-bp resolution (NCBI35/HG17 genome build).

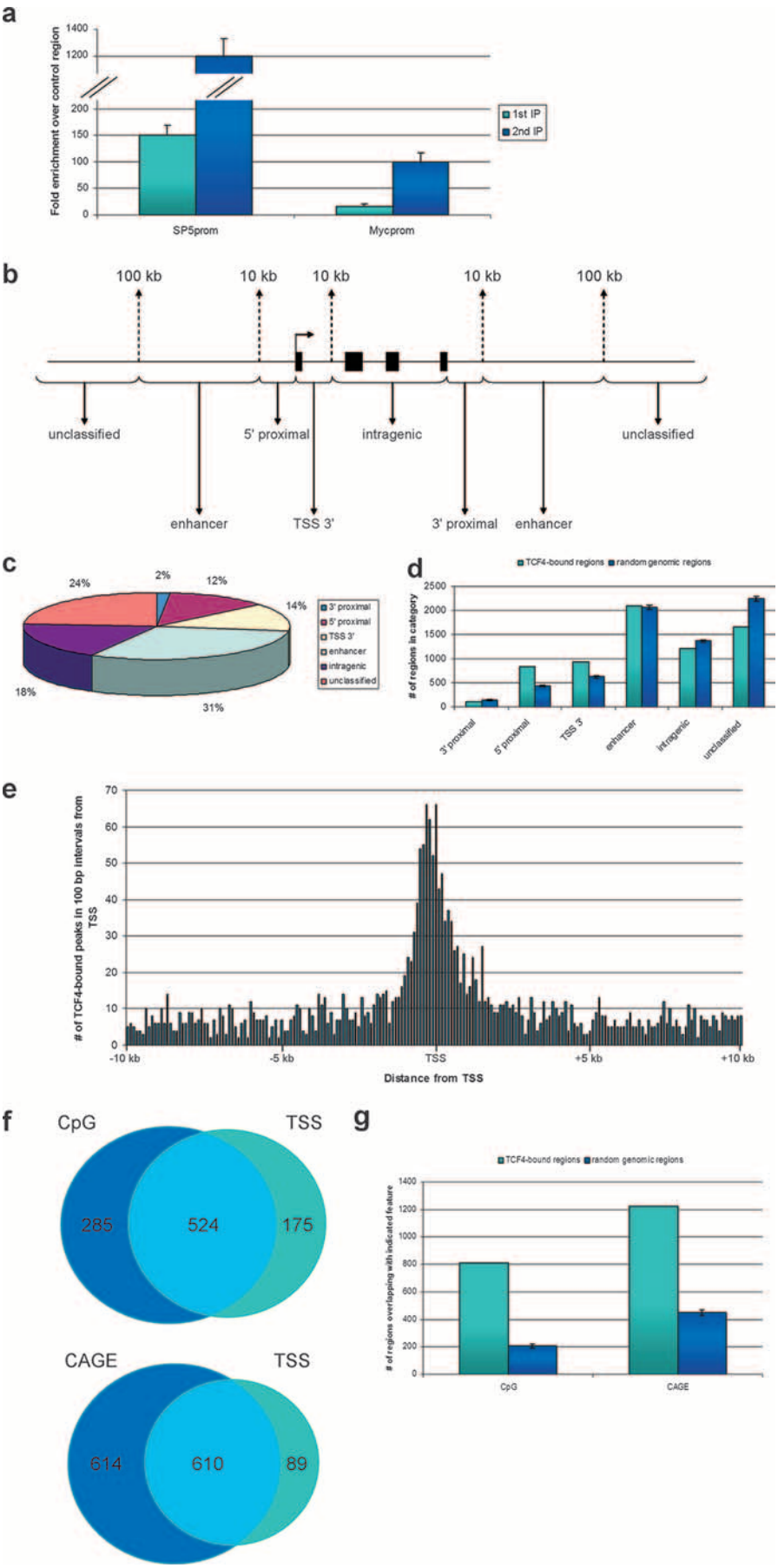
To verify the peaks obtained by the genome-wide array, two sets of triplicate experiments were performed, a dedicated array for chromosomes 1 to 22p and the tiling array for chromosomes 22q/X/Y. The dedicated array contained 1,251,695 oligonucleotides covering the putative TCF-4-bound sequences extracted from the genome-wide array (chromosomes 1 to 22p), plus a tiled region from chromosome 21 (chromosome 21: 33206900 to 46800000) at 100-bp resolution for normalization purposes. The dedicated array was divided over two slides, both containing the full tiled region. The replicates for the tiling array for chromosomes 22q/X/Y contained 769,784 oligonucleotides on two slides.

Identification of TCF-4-binding regions. Three different peak identification software packages were used to extract putative peaks from the genome-wide scans, Mpeak (MP) (http://www.stat.ucla.edu/~zmdl/mpeak), TileMap (TM) (21), and NimbleGen Peakdetection (NP) (NimbleGen, Inc.), to maximize the inclusion of putative TCF-4-binding regions on the dedicated array. MP (version 2.0) was used with default settings and a threshold of 2.5S.D.; TM was used with HMM (posterior probability of >0.5; maximal gap allowed, 100; UMS on; G0 p%, 0.01; G1 q%, 0.05; selection offset on; grid size, 1,000; expected hyblength, 50; no repeat filter; no test statistics) to combine neighboring probes. The Nimblegen program (version 2) was used with a 1% FPR cutoff. Identified peaks were extended 1,000 bp on either site from the center of the peaks, resulting in 67,838 peak areas. Probes for inclusion on the dedicated array were filtered using BLAT software (22), excluding probes aligning more than 10 times in the genome. Following three replicate hybridizations on the dedicated arrays and on the array covering chromosomes 22q, X, and Y, application of Tukey's biweight analysis on the chromosome 21 tile path was used to normalize and scale each slide (http://mathworld.wolfram.com/tukeysbiweight.html). The mean ratio signal and variance were calculated for each probe, and peak recognition with the same peak recognition algorithms as described above was performed with the mean ratio signal track. The gap parameter for both MP and TM was set to 250 bp, i.e., allowing a maximum of 250 bp between probes that constitute a peak. The rest of the parameter settings for the programs were adjusted to call approximately the same number of peaks with each method. Using a 2.5-standard-deviation cutoff for MP, a total of 15,282 or 1,176 peaks were called in the dedicated design or the chromosome 22q/X/Y set, respectively. Using these numbers as a reference, both NP and TM were tried iteratively with increasing or decreasing thresholds for peak detection to achieve a peak set of approximately 15,000 peaks in the dedicated design set and 1,100 peaks in the chromosome 22q/X/Y set. The final peak thresholds for NP were 0.14 (dedicated design) and 0.02 (chromosome 22q/X/Y). The final peak thresholds for TM were 0.10 (dedicated design) and 0.95 (chromosome 22q/X/Y). The overlap of peaks found with the different programs was determined by defining overlaps as peaks positioned within 1,000 bp of each other. The set of peaks found by TM which overlapped with both MP and NP peaks was chosen as the final peak set. The final peak set contains 11,912 peaks in the regions of the dedicated design set and 555 peaks in chromosome 22q/X/Y set (see Table S8 in the supplemental material). The final peak set was divided in four confidence groups by the mean signal and variance of the probes within a peak. A total of nine probes around the peak position were used to calculate the mean signal and variance for each peak. The peak confidence sets were divided around the median mean signal and the median variance of the dedicated array (set A, mean peak signal of >1.5 and mean peak variance of <0.5; set B, mean peak signal of >1.5 and mean peak variance of >0.5; set C, mean peak signal of <1.5 and mean peak variance of >0.5; set D, mean peak signal of <1.5 and mean peak variance of <0.5).

Comparison between TCF-bound region and random genomic regions. A randomization test was preformed in order to compare properties of TCF-bound regions with those of other genomic regions. One hundred or 250 (where indicated) random sets were sampled from the human genome assembly to retain the same region size and distribution between chromosomes as with the original 6,868 TCF-bound sites. All random peaks were chosen from the unmasked sequence that was interrogated by the ChIP-on-chip experiment. The analyses of TCF-bound region properties with respect to gene structure, CpG islands, capped analysis of gene expression (CAGE) tags, clustering of sites around TSS, presence of the TCF motif, and evolutionary conservation were performed for real and random sets.

Evolutionary conservation of TCF-bound regions and motifs. Pairwise nucleotide BlastZ-net human-mouse, human-rat, human-chicken, and human-dog alignments were taken from the Ensembl database (19). Total conservation at consensus TCF motifs, TCF-bound regions, and random regions (200 bp around the center of the peak in both cases) were calculated. Insertions/deletions and unaligned segments were excluded from this calculation.

Identification of transcription factor-binding sites in TCF4-binding regions. Matrices from the Transfac database (version 11.1) were searched for using the matrix scanning program Storm (37) with a per-match *P* value cutoff of 0.0001 and an Hg17 intergenic 8mer word table. The matches for each matrix were tabulated across the foreground (500 bp around peak centers) and background (1,000-bp flanking sequence around peak centers) sets. A proportion test was then performed using the statistical computing language R, specifically, the prop.test function of R version 2.6.1. To derive sequence logos from Transfac


a
Fold enrichment over control region
1400
1200
200
150
100
50
0
1st IP
2nd IP
SP5prom
Mycprom
b
100 kb
10 kb
10 kb
10 kb
100 kb
unclassified
5' proximal
intragenic
unclassified
enhancer
TSS 3'
3' proximal
enhancer
c
24%
2%
12%
14%
18%
31%
3' proximal
5' proximal
TSS 3'
enhancer
intragenic
unclassified
d
TCF4-bound regions
random genomic regions
# of regions in category
2500
2000
1500
1000
500
0
3' proximal
5' proximal
TSS 3'
enhancer
intragenic
unclassified
e
# of TCF4-bound peaks in 100 bp intervals from TSS
70
60
50
40
30
20
10
0
-10 kb
-5 kb
TSS
+5 kb
+10 kb
Distance from TSS
f
CpG
TSS
285
524
175
CAGE
TSS
614
610
89
g
TCF4-bound regions
random genomic regions
# of regions overlapping with indicated feature
1400
1200
1000
800
600
400
200
0
CpG
CAGE

matrices, a custom program was used. To generate logos from the Storm output, the WebLogo software program, version 2.8.2 (http://weblogo.berkeley.edu/), was used.

Biological function of TCF-bound genes. Genes upregulated in human primary adenomas and bound by TCF4 within 100 kb of their TSSs were interrogated for gene ontology category and KEGG (Kyoto encyclopedia of genes and genomes) pathway enrichment using the web-based tool g:Profiler (http://biit.cs.ee/gprofiler/) (34).

Microarray data accession numbers. The microarray data can be accessed at http://www.ebi.ac.uk/arrayexpress/, experiment code E-TABM-402.

RESULTS

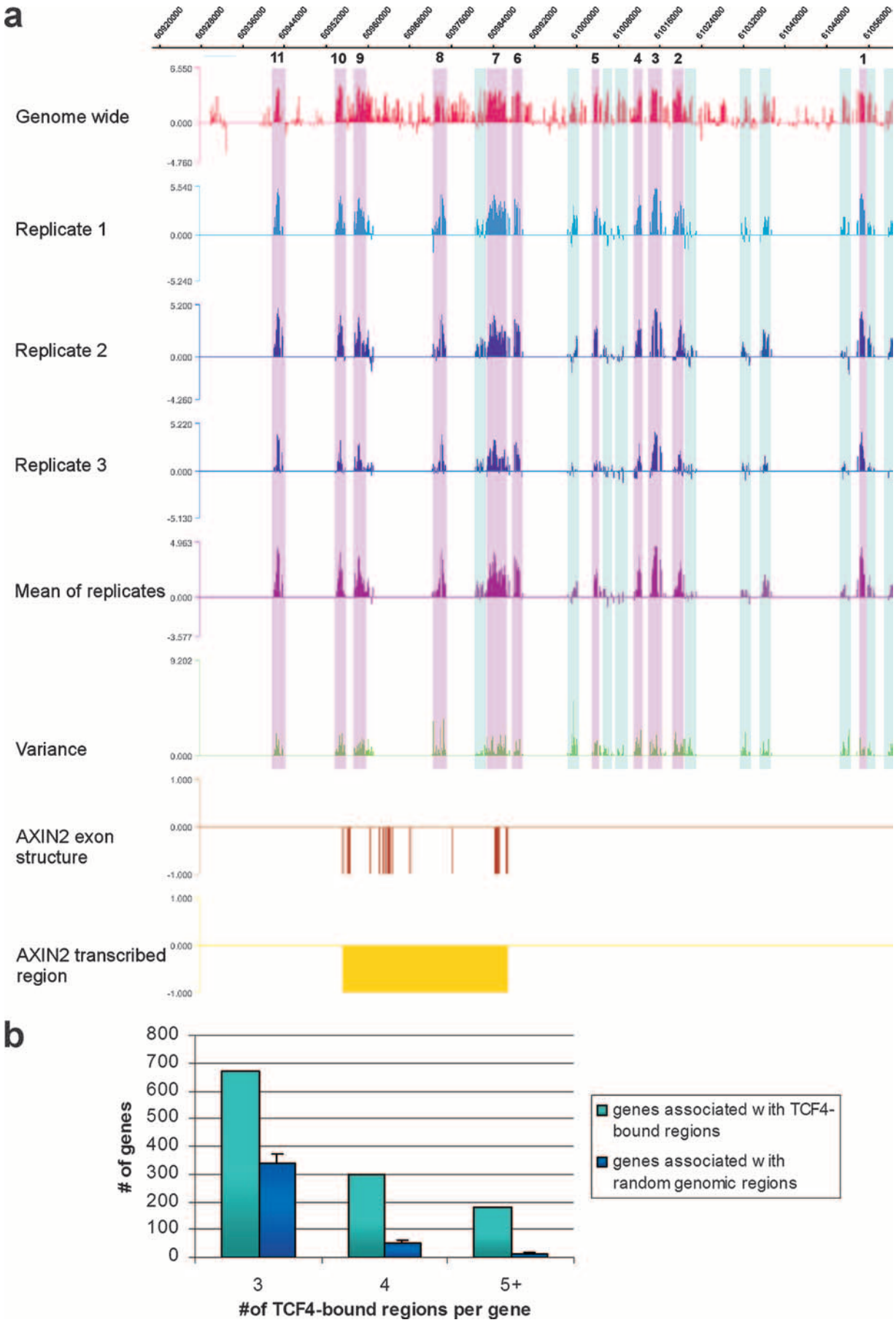
Genome-wide profile of TCF4 binding in CRC cells. To identify in vivo TCF4-binding sites in a comprehensive manner, we optimized sequential chromatin immunoprecipitations using a goat polyclonal antibody raised against the N terminus of the TCF4 protein. The increase in specific enrichment attained by the sequential immunoprecipitations (12) should allow the comprehensive identification of even "weak" TCF4-binding sites in the genome of CRC cells. All experiments were performed with the diploid, β-catenin mutant human colon cancer cell line LS174T, which expresses a TCF4-dependent transcriptional program similar to that which is physiologically driven by Tcf/β-catenin in the proliferative compartment of intestinal crypts (49). As shown in Fig. 1a (see also Fig. S1 in the supplemental material), the proximal promoter of the *SP5* gene, a previously described Wnt target (41, 43), was enriched >100-fold after one round of immunoprecipitation and was enriched >1,000-fold after two sequential immunoprecipitations using the anti-TCF4 antibody. Consistent with c-*Myc* being Wnt responsive in LS174T cells (49), the previously described TCF response element in the c-*Myc* promoter (17) was also enriched (Fig. 1a) (see Fig. S1 in the supplemental material), albeit to a lower extent. These observations were independently confirmed using an anti-TCF4 monoclonal antibody raised in our laboratory (1) (not shown).

DNA (either from input chromatin or from sequential ChIP material) was amplified by ligation-mediated PCR and labeled with Cy3 and Cy5, respectively. The probe samples were then hybridized to a set of 36 microarrays covering the repeat-masked regions of the human genome at 100-bp resolution (NimbleGen Systems, Inc.) (apart from the q arm of chromosome 22 and chromosomes X and Y; see below). Three different algorithms, MP (http://www.stat.ucla.edu/~zmdl/mpeak), TM (21), and NP (NimbleGen Systems, Inc.), were used to predict a total of 67,838 putative TCF4-binding sites. The application of all three programs redundantly aimed at the inclusion in the peak count of the greatest possible number of putative TCF4-binding sites and the minimization of false negatives. To verify the binding sites predicted from the genome-wide hybridization, we designed dedicated arrays covering regions of 2 kb around each detected peak (chromosomes 1 to 21 and 22p). ChIP-on-chip experiments were performed on the dedicated arrays with three biological replicates (independent TCF4 chromatin immunoprecipitates, independently amplified and labeled). The same replicates were used to probe in triplicate the 100-bp-resolution tiling path array covering the remaining chromosomes, 22q, X, and Y.

The peak detection procedure performed for both the replicates of the dedicated arrays and the replicates of the chromosome 22q/X/Y tiling path array was the following: The three biological replicates were merged into one data set by calculating the mean ratio signal for each probe. The three peak recognition algorithms were applied to the mean ratio signal track, and only peaks found by all three algorithms were retained to extract 11,912 binding regions from the dedicated arrays and 555 binding regions from the chromosome 22q/X/Y array. By requiring three out of three programs to detect each peak, we increased the stringency of peak prediction to minimize the inclusion of false positives in the final set of TCF4-binding sites. Prior to validation by quantitative PCR analysis, the detected peaks were further subdivided into four groups according to mean peak signal values and mean peak variance over a region of nine probes surrounding the peak center (set A, mean peak signal of >1.5 and mean peak variance of <0.5; set B, mean peak signal of >1.5 and mean peak variance of >0.5; set C, mean peak signal of <1.5 and mean peak variance of >0.5; set D, mean peak signal of <1.5 and mean peak variance of <0.5). For both the dedicated and chromosome 22q/X/Y binding sites, 15 randomly selected peaks from each of the 4 groups were validated by quantitative PCR. All 60 peaks from both sets A and B from the dedicated design, as well as chromosome 22q/X/Y, were positive. Only 8/15 and 6/15 peaks from set C and 7/15 and 9/15 peaks from set D for the dedicated design and chromosome 22q/X/Y, respectively, were positive in the qPCR assays (see Fig. S2 and S3 and Table S1 in the supplemental material). The accuracy rate for both the dedicated design and the chromosome 22q/X/Y sets of binding sites is 75%; this indicates that the three biological replicates on our dedicated design maintain the same specificity as the three biological replicates on the chromosome 22q/X/Y tiling array, validating the dedicated array approach, in agreement with other studies (23, 24).

Sets A and B gave an accuracy rate of 100%. Since sets C and D yielded accuracy rates between 40% and 60% and con-

FIG. 1. ChIPs over regions bound by TCF4 and genomic distribution of TCF4-bound regions. (a) Association of TCF4 with the proximal promoters of *SP5* and c-*Myc* was determined by single (light blue) or sequential (dark blue) ChIP followed by qPCR and expressed as relative enrichment over the nonbound exon 2 of the myoglobin gene. Error bars represent standard deviations for three independent experiments. (b) Schematic illustration delineating the criteria for binding-site classification with respect to a gene locus. (c) Localization of TCF4-binding sites in relation to annotation to nearest transcription units. Shown are percentages of binding sites in the different location categories as defined in panel b. (d) Distribution in categories, defined as in panel b, of TCF4-bound regions (light blue) or random genomic regions (dark blue). Error bars represent standard deviations of 100 random groups. (e) Distribution in 100-bp intervals of TCF4-bound regions located within 10 kb of annotated TSSs. (f) Venn diagrams depicting the number of TCF4-bound regions within 1 kb of CpG islands, annotated transcription start sites of protein-coding genes, or both (top) and the number of TCF4-bound regions within 1 kb of CAGE tags, annotated transcription start sites of protein-coding genes, or both (bottom). (g) Distribution, in categories defined as in panel f, of TCF4-bound regions (light blue) or random genomic regions (dark blue). Error bars represent standard deviations of 250 random groups.


a
Genome wide
Replicate 1
Replicate 2
Replicate 3
Mean of replicates
Variance
AXIN2 exon structure
AXIN2 transcribed region
11
10
9
8
7
6
5
4
3
2
1
b
# of genes
800
700
600
500
400
300
200
100
0
3
4
5+
#of TCF4-bound regions per gene
genes associated w ith TCF4-bound regions
genes associated w ith random genomic regions

tained peaks of mostly lower levels of specific enrichment than A and B, we continued our analyses with the binding regions of sets A and B only. Merging of peaks within 1,000 bp of each other in these two groups resulted in 6,868 high confidence TCF4-binding sites (see Table S2 in the supplemental material). We estimated that this approach may miss up to 2,150—mostly low-enrichment—binding sites but should increase the specificity of subsequent analyses.

As expected, the high-confidence peak set included prominent binding sites over the proximal promoters of the *SP5* and c-*Myc* genes (not shown). An additional 44 TCF4-binding sites from peak sets A and B near known target genes of the pathway (45, 48, 49) were all confirmed by qPCR (see Fig. S4 and Table S3 in the supplemental material), further underscoring the specificity of the generated TCF4-binding profile.

We also proceeded to investigate the presence of validated TCF4-binding sites in other CRC cell lines. To this effect, chromatin immunoprecipitations with the goat polyclonal antibody against TCF4 were performed with HCT116 and DLD1 cells, and 25 randomly selected binding sites were tested by qPCR (see Fig. S5 in the supplemental material). Of the 25 tested binding regions, 20 (80%) were positive in HCT116 cells and 24 (96%) were positive in DLD1 cells. The high percentage of TCF4-binding sites bound in all three cell lines further stresses the relevance of the generated TCF4-binding profile for the investigation of TCF4-mediated transcriptional regulation in CRC.

Distribution of TCF4-binding sites with respect to gene structure. To evaluate the distribution of the TCF4-binding sites along the genome, we annotated these with respect to the TSS of the nearest gene (based on Ensembl v34 (6). Peaks were defined as either 5′-proximal (10 kb upstream of the TSS), TSS 3′ (10 kb downstream of the TSS), intragenic (within gene bodies, from 10 kb 3′ from the TSS to the gene end), 3′ proximal (within 10 kb downstream of the gene), or distal "enhancer" (10 to 100 kb either up- or downstream of gene boundaries). Peaks located more than 100 kb away from the nearest gene were annotated as unclassified (Fig. 1b).

Eight hundred thirty-nine (12%) of peaks were found within 5′-proximal locations, 941 (14%) were located in TSS 3′ positions, and 117 (2%) within 3′-proximal locations. One thousand two hundred nine (18%) peaks were found within genes, further than 10 kb from the TSS. Two thousand ninety-eight (31%) peaks were located in putative long-range "enhancer" positions (up to 100 kb up- or downstream of a gene). One thousand six hundred sixty-four (24%) peaks were not located within 100 kb of the boundaries of the nearest gene (unclassified) (Fig. 1c). When this distribution of peaks was compared to that of random genomic fragments, it became apparent that there was a striking bias for TCF4-binding sites within 10 kb both up- and downstream of TSSs (Fig. 1d). The pronounced clustering of TCF4-bound regions around TSSs can be prominently observed in Fig. 1e, a plot of the distribution of binding sites relative to the distance from the TSS. Despite this conspicuous pattern observed for peaks near TSS, more than 70% of TCF4-bound regions are located at distances greater than 10 kb from the nearest annotated transcription starts, a distribution which is similar to that determined using similar global approaches for other sequence-specific DNA-binding transcription factors, such as Oct4 and Nanog (29), p53 (51), and ER (8).

We also analyzed the overlap of TCF4-bound regions with respect to CpG islands and found 809 of them to be within 1,000 bp of annotated CpG islands (Fig. 1f), a number much greater than that observed for random genomic regions (Fig. 1g). Significantly, 285 (35%) of the TCF4-bound regions overlapping CpG islands were not in similar proximity (within 1 kb) to TSSs of protein coding genes (Fig. 1f).

Visual inspection of the distribution of the TCF4-binding regions revealed another interesting observation: peaks frequently cluster around putative target genes. An extreme example was provided by *AXIN2*, a well-known target gene of the Wnt pathway (31), which associates with no fewer than 11 peaks within 100 kb of its TSS (Fig. 2a). We explored whether this clustered distribution of peaks around genes was nonrandom by comparing it to the distribution expected for randomly selected genomic regions. The analysis shown in Fig. 2b demonstrates that the distribution was indeed not random, since there were significantly more genes that associate with three or more TCF4-binding sites than expected, providing statistical validation to this striking phenomenon.

Determination and conservation of the TCF4-binding DNA motif. In in vitro selection-based assays, we have previously defined the optimal TCF-binding motif as AAGATCAAAGG (44, 46). Using a different in vitro approach, Hallikas and colleagues defined a slightly shorter optimal TCF4 binding motif: CATCAAAGG (14). We proceeded to mine the underlying sequence of the TCF4-bound peaks to determine the *cis* element(s) which mediates TCF4 binding in vivo. We applied MDscan, a de novo motif discovery algorithm (28), using random samples of the peaks validated by qPCR for program training. The most common motifs discovered within windows of different lengths bore a strong resemblance to the consensus motif identified in the in vitro studies. Three examples of these—the most common motif within a 7-bp, 11-bp, or 15-bp window—are depicted in Fig. 3a. Seventy percent (4,793/6,868) of the sites bound by TCF4 contained at least the shortest (7 nucleotides) motif uncovered by our method, and this ratio was much greater than that expected for random genomic fragments. This statistical significance held true also for the occurrence of the longer motifs (Fig. 3b). We further examined the evolutionary conservation of the TCF4-binding regions and DNA-binding motifs with respect to the genomes of rat and mouse (Fig. 3c), as well as dog and chicken (not shown): Both

FIG. 2. TCF4-binding-site clustering around target genes. (a) TCF4-binding-site distribution around the *AXIN2* gene. Depicted is the binding pattern of TCF4 around *AXIN2* as revealed by the genome-wide experiment and the three independent biological replicates on the dedicated array, including the mean and variance tracks from the three replicates. High-confidence peaks are highlighted in magenta and numbered 1 to 11, low-confidence peaks in light blue. (b) Numbers of genes bound within 100 kb of their TSSs by three, four, or five or more TCF4-binding sites (light blue) or random genomic regions (dark blue). Error bars represent standard deviations of 100 random groups.

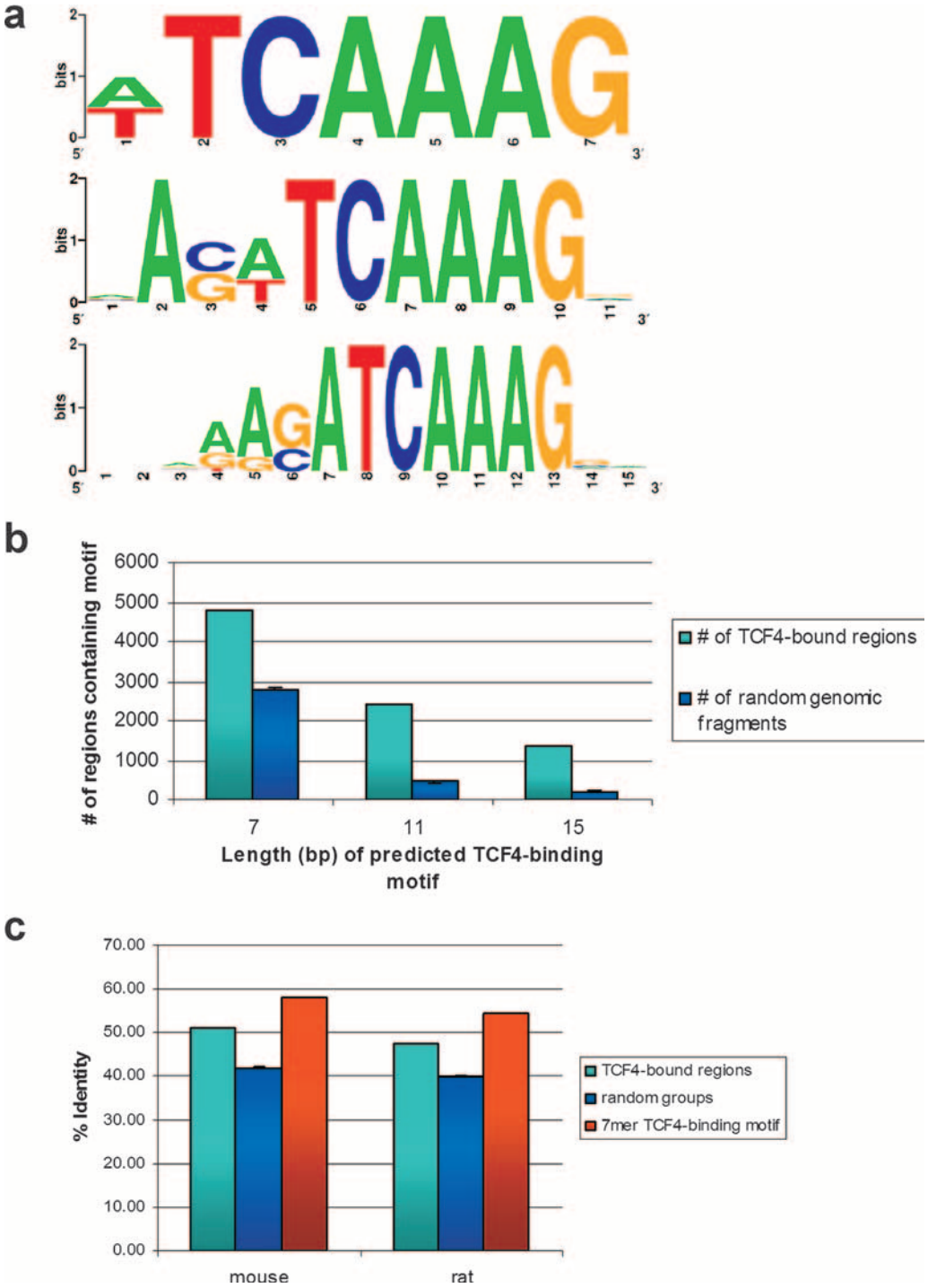


FIG. 3. (a) Sequence logos illustrating the nucleotide distribution for the in vivo TCF4 consensus sites of 7, 11, and 15 bp, as defined by MDscan. (b) Number of TCF4-bound (light blue) or random genomic (dark blue) regions containing the indicated TCF4-binding motif, as depicted in panel a. Error bars represent standard deviations of 100 random groups. (c) Percent identities of TCF4-bound regions (light blue), random genomic regions (dark blue), and the 7-mer TCF4-binding motif (red, as depicted in panel a) for mouse-human and rat-human pairwise genomic comparisons. Error bars represent standard deviations of 100 random groups.

the sequences surrounding the centers of the peaks and the TCF4-binding motifs contained within the sequences were significantly more conserved compared to random genomic segments, as expected for functional transcriptional regulatory regions (chi-square, $P < 0.01$). These observations further underscored the validity of our TCF4-binding sites.

We further mined the sequences encompassing the TCF4-binding regions to identify binding sites for other transcription

TABLE 1. Transfac matrices enriched around TCF4 binding-site centers[a]

Binding factor (accession no.)	Transfac-derived sequence logo used in search	Sequence logo of motifs actually present in TCF4 binding region peak centers	*P* value for enrichment relative to flanking regions	% of TCF4 binding sites containing at least 1 instance of motif
NF-1 (M00806)			1.31E−53	21.8
HNF4 (M00411)			1.46E−50	21.4
AP-1 (M00199)			3.62E−48	14.6
PPARγ (M00512)			7.77E−33	14.6
Elk-1 (M00025)			5.46E−31	19.8
GATA-3 (M00351)			9.38E−26	13.1
c-Ets-1 (M00032)			3.16E−24	16.7
FoxD1 (M00292)			4.97E−22	13.5
Bach1 (M00495)			5.16E−19	16.4

[a] Depicted are sequence logos of transcription factor binding sites identified as significantly enriched when sequences surrounding TCF4 binding-site centers (500 bp around each center) are compared to the 1,000-bp regions flanking the binding-site centers on either side.

factors using the Transfac database. Matrices were searched for using the program Storm (37) in the 500-bp genomic regions surrounding the TCF4-binding-region centers, and their incidence was compared to incidence in the 1,000-bp genomic regions flanking the peak centers on either side. This analysis revealed the presence of a number of motifs for transcription factors, such as NF1, HNF4, PPARγ, and others, specifically enriched in TCF4-binding regions (Table 1). These factors potentially coregulate transcription of TCF4 target genes.

Correlation between TCF4 occupancy and Wnt-dependent transcriptional regulation. We have previously described the global transcriptional program driven by the Wnt pathway in CRC (36, 45, 48, 49). In references 36 and 45, we and collaborators performed an exhaustive array-based comparison of the Wnt target gene program with colorectal cell lines and primary human adenomas. These expression data sets were used to investigate the potential correlation between Wnt-mediated transcriptional effects and the genome-wide TCF4 binding profile. A stepwise differential expression rank analysis showed a significant correlation between TCF4 occupancy of target-gene regulatory regions and genes upregulated in human primary adenomas compared to normal colonic mucosa; genes with TCF4-binding regions within 100 kb of their TSS were more likely to show significantly upregulated expression than genes without (Fig. 4). Significant correlation between TCF4 occupancy and expression profile changes was also observed in LS174T cells inducibly overexpressing an N-terminally truncated dominant-negative mutant form of TCF4 (ΔNTCF4) (see Fig. S6 in the supplemental material). These

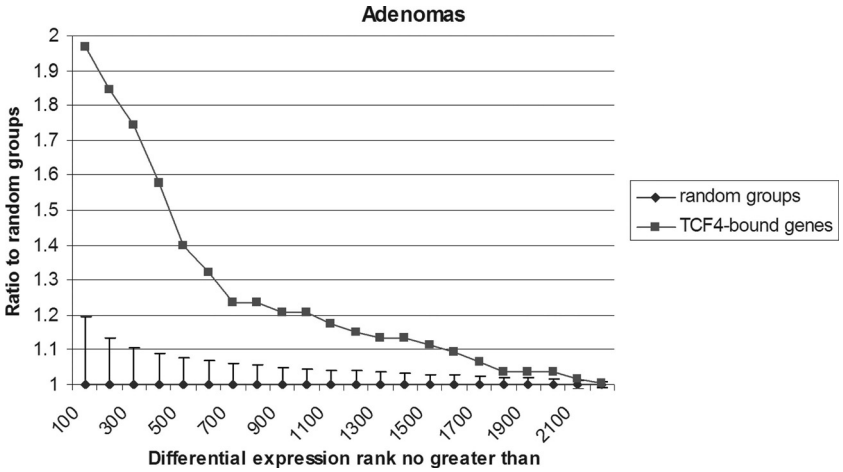


FIG. 4. Correlation of TCF4 binding and TCF4/β-catenin-controlled gene expression. Differential expression rank analysis for genes bound within 100 kb of TSS by TCF4 or random groups from genes upregulated in human primary adenomas, using a step size of 100. Error bars represent standard deviations of 100 random groups.

data demonstrate that Wnt-dependent transcriptional changes correlate strongly with direct TCF4 occupancy of regulatory regions, even when the sources of the binding and expression profiles are different (CRC cell lines versus primary adenomas).

To examine the possibility that TCF4-binding sites have the potential to regulate RNA species not profiled by the above expression microarray experiments, we overlapped our TCF4-bound regions with CAGE tags, generated by the FANTOM3 (functional annotation of mouse 3) consortium (http://fantom.gsc.riken.go.jp/) (7), to define often previously unknown TSSs. We found that 1,224 TCF4-bound regions were within 1 kb and 3,324 were within 10 kb of human CAGE tags, a colocalization much greater than expected for random genomic regions (Fig. 1f and g; also data not shown). Significantly, 614 of the 1,224 (50%) and 1,729 of the 3,224 (54%) of TCF4-bound regions overlapping CAGE tags within 1 and 10 kb, respectively, did not overlap TSSs of known protein coding genes within the same distances (Fig. 1f; also data not shown). This provides an indication that many TCF4-bound regions may regulate transcription of novel RNA species not profiled by conventional expression microarrays.

Biological functions of TCF4 target genes. Functional categorization of TCF4 target genes (genes upregulated in human primary adenomas and bound by TCF4 within 100 kb of TSSs) revealed enrichment of genes involved in a broad spectrum of functions, such as cell proliferation ($P = 4.34 \times 10^{-9}$), transcription ($P = 5.3 \times 10^{-7}$), cell adhesion ($P = 6.19 \times 10^{-6}$), and the proteasome complex ($P = 5.09 \times 10^{-8}$) (see Table S4 in the supplemental material). Further examination of genes bound by TCF4 within 10 kb of TSS (irrespective of whether they were upregulated in human adenomas) revealed additional enriched categories, including negative regulation of programmed cell death ($P = 9.6 \times 10^{-6}$) and establishment and maintenance of chromatin ($P = 7.7 \times 10^{-7}$) (see Table S4 in the supplemental material). Promotion of cell proliferation and the negative regulation of apoptosis are functions consistent with the activity of a transcription factor at the end point of the Wnt pathway, which is involved in maintaining the proliferative compartment of the mammalian intestinal crypt and in carcinogenesis. The list of bound genes also contains a large number of sequence-specific transcription factors, many of which were not previously known to be targets of the Wnt signaling pathway. The abundance of sequence-specific transcription factors among the TCF4-bound genes should clarify regulatory relationships that will help distinguish direct from indirect targets of the pathway. It is noteworthy that these targets include three members of the TCF family, LEF1, TCF7 (TCF1), and TCF7L2 (TCF4) itself. It should further be noted that KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways with components enriched in the TCF4-bound gene set included the Wnt pathway itself ($P = 7.7 \times 10^{-6}$) and axon guidance ($P = 8.9 \times 10^{-6}$) (see Table S4 in the supplemental material). The latter contains the previously identified targets EPHB2 and EPHB3 (3, 4), which serve to position cells in the intestinal epithelium along the crypt/villus axis. Other genes in this category may be also involved in similar processes.

Transcriptional regulatory activity of TCF4-bound regions. We next investigated whether the identified TCF4-bound genomic regions exert transcriptional regulatory activity. Fragments of approximately 1,000 bp surrounding 22 peaks (see Table S5 in the supplemental material) were cloned either as promoters (in the case of peaks that were located in the vicinity of the TSSs of target genes) or as enhancers upstream of a minimal fragment encompassing the TATA box of the adenovirus major late promoter. The resulting plasmids were transiently transfected into LS174T cells. Ten of the 22 regions enhanced transcription of the luciferase reporter in this assay. These included the proximal promoter of *SP5*, a region far downstream of the *ADRA2C* gene, which was the strongest enhancer tested at more than 90-fold the activity of the control, as well as the 3′ and intronic peaks associated with the *BMP7* gene. Cotransfection of ΔNTCF4 led to downregulation of the activity of nine elements (Fig. 5a). As a control experiment, we cloned 15 random genomic regions as promoters and 15 random genomic regions as enhancers in front of the same luciferase reporter. Of these, only three were transcriptionally ac-

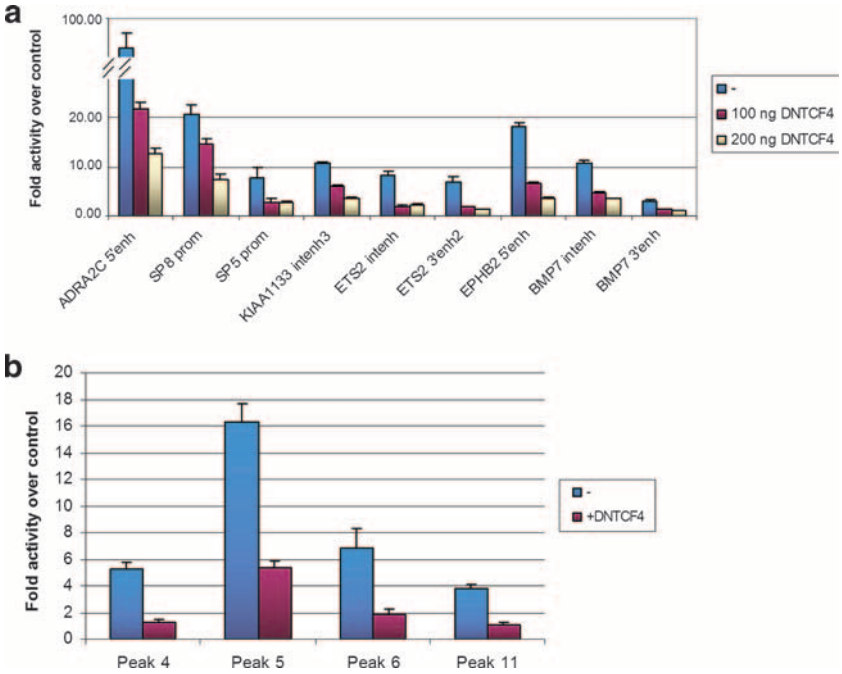


FIG. 5. Transcriptional activity of TCF4-bound regions in CRC cells. (a) TCF4-binding regions were cloned into the pGL3b or pGL3/AdMLTATA vector, in the case of TSS-proximal or non-TSS-proximal regions, respectively, and transfected into Ls174T with cotransfection of the CMV-*Renilla* vector as the normalizing control and with or without cotransfection of ΔNTCF4. Values are expressed as activity relative to that of the respective empty pGL3 vectors. Error bars represent standard deviations for three independent experiments. (b) Eleven TCF4-binding regions surrounding the *AXIN2* gene within 100 kb of the TSS were cloned into the pGL4.10 or pGL4.10/TATA vector, in the case of TSS-proximal or non-TSS-proximal regions, respectively, and transfected into LS174T/ΔNTCF4 cells with cotransfection of the CMV-*Renilla* vector as the normalizing control and with or without doxycycline treatment to induce ΔNTCF4 expression. Values are expressed as activity relative to that of the respective empty pGL4.10 vectors. Error bars represent standard deviations for three independent experiments.

tive and none was regulated by cotransfection of ΔNTCF4 (data not shown).

We further investigated the transcriptional activity of all peaks surrounding the *AXIN2* gene. The latter is a well-known target of the pathway (31) and is surrounded by at least 11 TCF4-binding regions within 100 kb of its TSS (Fig. 2a) (see Table S5 in the supplemental material). Using a similar strategy to that described above, we found that of those 11 peaks, 4 enhanced the transcriptional activity of a luciferase reporter and were downregulated by overexpression of ΔNTCF4 (Fig. 5b). The regions active in these experiments include a peak near the TSS of the gene (peak 6) in a region previously shown to display Wnt-regulated transcriptional activity (20), two peaks 5′ of the TSS, and 1 peak 3′ to the end of the transcription unit.

These experiments demonstrated that a significant subset of TCF4-bound regions uncovered by the ChIP-on-chip approach score as Wnt-responsive transcriptional regulatory regions in transient reporter gene assays. The subset of Wnt-regulated regions included both peaks near TSSs (i.e., the SP5, SP8, and AXIN2 proximal promoters and an EPHB2 5′ peak) and binding regions further away from TSS (i.e., an ADRA2C 3′ peak, the AXIN2 far-upstream peaks, and the BMP7 and ETS2 intronic peaks), consistent with the Wnt pathway having the ability to regulate transcription of target genes from large distances.

DISCUSSION

Genome-wide approaches for the identification of transcription factor-binding sites are increasingly becoming the tools of choice for the elucidation of the transcriptional circuitries governing development, homeostasis, stem cell biology, or the genesis of cancer (9–11, 29, 51). The ChIP-on-chip approach we have employed here has allowed us to comprehensively map genome-wide chromatin occupancy by TCF4, the transcription factor at the end point of Wnt signaling in the mammalian intestine. Our experiments reveal that the binding profile of TCF4 resembles that of other studied transcription factors, such as ER and p53 (29, 51), in that TCF4 binding is observed both in the vicinity and also at sites located at great distances from the TSSs of annotated genes. An interesting and novel observation to emerge from our results is that TCF4-binding sites frequently cluster in the vicinity of putative target genes. *AXIN2* is a characteristic example, with TCF4-binding sites located in intronic, far-upstream, and downstream locations. This multitude of binding sites around genes like *AXIN2* may serve multiple regulatory purposes; while as demonstrated in this study, some of the binding sites identified act as classical transcriptional regulatory elements, including regions both upstream and downstream of the gene, the multiple TCF4-binding regions around *AXIN2* may serve other purposes, such as maintaining an open chromatin domain or providing a more

accurate sensor for the intranuclear β-catenin concentration. It should in the least provide a source of mechanistic insight in future studies of Wnt-dependent transcriptional regulation.

Our study correlates the global profile of TCF4 binding with differential expression array-based data to provide a view of the direct targets of the Wnt pathway in the mammalian intestine. The correlation between the primary adenoma-derived expression data and the cell-line-derived TCF4-binding pattern is particularly striking in that the two data sets are derived from different, albeit both Wnt-driven, sources. It should be noted here that only 12.5% (282/2,248) of the genes upregulated in adenomas were bound by TCF4 within 10 kb and only 20.5% (462/2,248) were bound within 100 kb of the transcription start, the limit of annotation applied to these analyses. Many indirect targets are likely to exist in the upregulated genes, since a number of genes bound by TCF4 encode transcription factors themselves, as well as more-direct targets, with TCF4-binding sites further away from the TSS. Conversely, only 12.5% (462/3,676) of the genes bound by TCF4 within 100 kb of the transcription start site were significantly upregulated in adenomas. This is in line with what has been reported in previous studies (52, 54) and most likely has both technical and biological reasons: slight expression level changes below the limit of detection of these analyses may contribute to the underdetection of valid TCF4 targets. Furthermore, functional redundancy in enhancer and transcription factor action may contribute to the lack of detectable transcriptional changes at some TCF4-occupied genes. Additionally, TCF4-binding sites located at greater distances from transcription start sites and annotated to the closest gene may in fact be exerting their regulatory function elsewhere, including on other genes further away or even on other chromosomes (40) or on noncoding regulatory RNAs not profiled in these studies; the last is also suggested by the significant overlap between TCF4-binding sites and CAGE tags.

Our approach has also allowed us to use the sequence underlying the TCF4 peaks to determine the in vivo TCF4-binding motif. The motif thus generated is very similar to motifs determined through in vitro experiments. Moreover, the motif is statistically overrepresented in the TCF4 peaks compared to occurrence in random genomic fragments, as expected for functional TCF4-binding sites, and both the TCF4-binding motifs and the underlying sequence of the TCF4-bound regions are evolutionarily conserved. It should be noted that some TCF4-binding regions do not contain a recognizable TCF motif (2,075/6,868; 30%). TCF4 may be recruited to these sites by an atypical binding motif not identified by our analyses or through protein-protein interactions with other factors directly recruited to these regions. More likely, TCF4 association with these sites may be indirect, mediated by enhancer "looping" effects: recruitment may be mediated by physical association of distinct genomic regions in *cis* looping out the intervening DNA (15, 16, 39) or between regions located on other chromosomes (30, 40). Additional experiments are under way to distinguish between these possibilities.

In a previously published study, the Enhancer Element Locator (EEL) computational tool developed by Hallikas and colleagues integrated conservation of in vitro-determined binding sites along with affinity and clustering information to predict TCF4-controlled enhancers (14). EEL predicted 130 putative Wnt-responsive enhancers containing 2 or more TCF4-binding sites, only 10 of which overlap (are within 1,000 bp of each other) with our experimentally validated set of 6,868 peaks. This overlap is slightly greater than random coincidence would allow (see Fig. S7 and Table S6 in the supplemental material). In order to exclude the possibility that the limited overlap between our data sets was caused by a failure of our ChIP-on-chip approach to uncover these binding sites, 10 randomly selected EEL-predicted enhancers (see Table S6 in the supplemental material) were tested by quantitative PCR on TCF4-ChIP material from LS174T cells. All sites tested were negative (enriched $<$2-fold over a control region in qPCR assays; data not shown), excluding the possibility that EEL-predicted enhancers are missed as false negatives. This means that the EEL bioinformatics tool predicts $<$0.15% of sites occupied by TCF4 in CRC cells, despite the significantly higher-than-random sequence conservation of our peaks. Of course, it is not unlikely that some of the remaining predicted enhancers not occupied in our CRC cells may represent authentic Wnt-responsive regulatory elements in other contexts. Comparison of the two studies does, however, underscore the fact that current computational tools are limited in their ability to predict the full complement of sites occupied by a transcription factor in a tissue of interest.

While this article was in preparation, a study was published identifying β-catenin-binding sites in the human CRC cell line HCT116, using serial analysis of chromatin occupancy (53). Of the 412 binding sites identified by Yochum et al., 293 binding sites are represented on the NimbleGen genome-wide arrays used in this study and are possible candidates for overlap with the TCF4-binding sites identified here. Of those 293 β-catenin-binding sites, 52 (18%) overlapped with our 6,868 TCF4-binding regions, a proportion which, albeit relatively small, was much greater than that determined for random genomic sequences (see Fig. S8 and Table S7 in the supplemental material). The overlap calculated for the 252 β-catenin-binding sites that contained a consensus TCF4-binding motif within 5 kb and the 4,793 TCF4-binding regions containing $\geq$1 TCF4 motif within 1 kb was similar (38 binding regions; 16%) and still significant (see Fig. S8 in the supplemental material). The incomplete overlap between the two sets of locational information may be due to the different experimental approaches (ChIP-on-chip versus serial analysis of chromatin occupancy, immunoprecipitations against TCF4 versus β-catenin, respectively).

A number of TCF4-binding regions act as Wnt-responsive promoters or enhancers in transient-transfection experiments, including regions both in the vicinity of and at great distances from transcription start sites. However, more than half (20/33) of TCF4-bound regions were inactive or nonregulated in this assay. Some regions may exert their regulatory activity through effects on the surrounding chromatin template, effects that may be difficult to recapitulate on transiently transfected templates. In the case of the 5′ hypersensitive sites of the β-globin locus control region, the enhancer activity of only 5′ HS2 is detectable in transient-transfection experiments whereas that of HS3 and -4 only becomes apparent when these are integrated into chromatin (27). In this respect, the binding of TCF4 may serve to regulate histone modifications and/or chromatin structure over these regions, since it has been demonstrated to interact

through β-catenin both with chromatin remodelers, such as Brg1 (2), and with the histone modifiers MLL and p300/CBP (18, 38, 42). Interestingly, TCFs have also been shown to exert potent intrinsic DNA-bending activity (13, 47, 50). These actions, rather than impinging directly on preinitiation complex formation on promoters of regulated genes, may serve a chromatin opening function, maintaining chromatin domains in a "poised" conformation and facilitating subsequent events involved in transcriptional activation. This model would be compatible with the multiplicity of sites, only some of which act as classical transcriptional regulatory elements, surrounding some target genes, such as *AXIN2*. Intriguingly, these potential activities of the TCF4/β-catenin complex might be modulated—facilitated or repressed—by other transcription factors which may bind with them on the same genomic regions, as predicted by the enrichment of the TCF4-binding regions in relevant transcription factor-binding matrices.

In conclusion, the current study provides a genome-wide binding profile of TCF4, the major transcription factor at the end point of Wnt signaling in the intestine. Combination of this locational information and differential expression data allows the delineation of the direct transcriptional targets of TCF4 in the human intestine and unveils Wnt-responsive *cis* elements by which their expression is controlled.

ACKNOWLEDGMENTS

We thank members of the Clevers and Stunnenberg laboratories for help and discussions, Tokameh Mahmoudi for critical reading of the manuscript, Thanasis Margaritis for assistance with bioinformatics. and Andrea Haegebarth for help with figure preparation.

P.H. is supported by successive European Molecular Biology Organization and Human Frontier Science Program Organization long-term fellowships. M.A.V.D. is supported by EU-FP6 IP EPITRON and STREP X-TRA-NET. S.D. is supported by EU-FP6 IP HEROIC.

REFERENCES

1. **Barker, N., G. Huls, V. Korinek, and H. Clevers.** 1999. Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium. Am. J. Pathol. **154:**29–35.
2. **Barker, N., A. Hurlstone, H. Musisi, A. Miles, M. Bienz, and H. Clevers.** 2001. The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation. EMBO J. **20:**4935–4943.
3. **Batlle, E., J. Bacani, H. Begthel, S. Jonkheer, A. Gregorieff, M. van de Born, N. Malats, E. Sancho, E. Boon, T. Pawson, S. Gallinger, S. Pals, and H. Clevers.** 2005. EphB receptor activity suppresses colorectal cancer progression. Nature **435:**1126–1130.
4. **Batlle, E., J. T. Henderson, H. Beghtel, M. M. van den Born, E. Sancho, G. Huls, J. Meeldijk, J. Robertson, M. van de Wetering, T. Pawson, and H. Clevers.** 2002. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell **111:** 251–263.
5. **Bernstein, B. E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, E. J. Kulbokas III, T. R. Gingeras, S. L. Schreiber, and E. S. Lander.** 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell **120:**169–181.
6. **Birney, E., D. Andrews, M. Caccamo, Y. Chen, L. Clarke, G. Coates, T. Cox, F. Cunningham, V. Curwen, T. Cutts, T. Down, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, S. Graf, M. Hammond, J. Herrero, K. Howe, V. Iyer, K. Jekosch, A. Kahari, A. Kasprzyk, D. Keefe, F. Kokocinski, E. Kulesha, D. London, I. Longden, C. Melsopp, P. Meidl, B. Overduin, A. Parker, G. Proctor, A. Prlic, M. Rae, D. Rios, S. Redmond, M. Schuster, I. Sealy, S. Searle, J. Severin, G. Slater, D. Smedley, J. Smith, A. Stabenau, J. Stalker, S. Trevanion, A. Ureta-Vidal, J. Vogel, S. White, C. Woodwark, and T. J. Hubbard.** 2006. Ensembl 2006. Nucleic Acids Res. **34:**D556–D561.
7. **Carninci, P., A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A. Semple, M. S. Taylor, P. G. Engstrom, M. C. Frith, A. R. Forrest, W. B. Alkema, S. L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S. M. Grimmond, C. A. Wells, V. Orlando, C. Wahlestedt, E. T. Liu, M. Harbers, J. Kawai, V. B. Bajic, D. A. Hume, and Y. Hayashizaki.** 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. **38:**626–635.
8. **Carroll, J. S., X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoute, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver, and M. Brown.** 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell **122:**33–43.
9. **Carroll, J. S., C. A. Meyer, J. Song, W. Li, T. R. Geistlinger, J. Eeckhoute, A. S. Brodsky, E. K. Keeton, K. C. Fertuck, G. F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E. A. Fox, P. A. Silver, T. R. Gingeras, X. S. Liu, and M. Brown.** 2006. Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. **38:**1289–1297.
10. **Cawley, S., S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl, and T. R. Gingeras.** 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell **116:**499–509.
11. **Cheng, A. S., V. X. Jin, M. Fan, L. T. Smith, S. Liyanarachchi, P. S. Yan, Y. W. Leu, M. W. Chan, C. Plass, K. P. Nephew, R. V. Davuluri, and T. H. Huang.** 2006. Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. Mol. Cell **21:**393–404.
12. **Denissov, S., M. van Driel, R. Voit, M. Hekkelman, T. Hulsen, N. Hernandez, I. Grummt, R. Wehrens, and H. Stunnenberg.** 2007. Identification of novel functional TBP-binding sites and general factor repertoires. EMBO J. **26:**944–954.
13. **Giese, K., J. Cox, and R. Grosschedl.** 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell **69:**185–195.
14. **Hallikas, O., K. Palin, N. Sinjushina, R. Rautiainen, J. Partanen, E. Ukkonen, and J. Taipale.** 2006. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell **124:**47–59.
15. **Hatzis, P., I. Kyrmizi, and I. Talianidis.** 2006. Mitogen-activated protein kinase-mediated disruption of enhancer-promoter communication inhibits hepatocyte nuclear factor 4α expression. Mol. Cell. Biol. **26:**7017–7029.
16. **Hatzis, P., and I. Talianidis.** 2002. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol. Cell **10:**1467–1477.
17. **He, T. C., A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein, and K. W. Kinzler.** 1998. Identification of c-MYC as a target of the APC pathway. Science **281:**1509–1512.
18. **Hecht, A., K. Vleminckx, M. P. Stemmler, F. van Roy, and R. Kemler.** 2000. The p300/CBP acetyltransferases function as transcriptional coactivators of beta-catenin in vertebrates. EMBO J. **19:**1839–1850.
19. **Hubbard, T. J., B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney.** 2007. Ensembl 2007. Nucleic Acids Res. **35:**D610–D617.
20. **Jho, E. H., T. Zhang, C. Domon, C. K. Joo, J. N. Freund, and F. Costantini.** 2002. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol. Cell. Biol. **22:**1172–1183.
21. **Ji, H., and W. H. Wong.** 2005. TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics **21:**3629–3636.
22. **Kent, W. J.** 2002. BLAT—the BLAST-like alignment tool. Genome Res. **12:**656–664.
23. **Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov, R. D. Green, M. Q. Zhang, V. V. Lobanenkov, and B. Ren.** 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell **128:**1231–1245.
24. **Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green, and B. Ren.** 2005. A high-resolution map of active promoters in the human genome. Nature **436:**876–880.
25. **Kinzler, K. W., and B. Vogelstein.** 1996. Lessons from hereditary colorectal cancer. Cell **87:**159–170.
26. **Korinek, V., N. Barker, P. J. Morin, D. van Wichen, R. de Weger, K. W. Kinzler, B. Vogelstein, and H. Clevers.** 1997. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC−/− colon carcinoma. Science **275:**1784–1787.
27. **Li, Q., S. Harju, and K. R. Peterson.** 1999. Locus control regions: coming of age at a decade plus. Trends Genet. **15:**403–408.
28. **Li, W., C. A. Meyer, and X. S. Liu.** 2005. A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. Bioinformatics **21**(Suppl. 1)**:**i274–i282.

29. **Loh, Y. H., Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim, and H. H. Ng.** 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. **38:**431–440.
30. **Lomvardas, S., G. Barnea, D. J. Pisapia, M. Mendelsohn, J. Kirkland, and R. Axel.** 2006. Interchromosomal interactions and olfactory receptor choice. Cell **126:**403–413.
31. **Lustig, B., B. Jerchow, M. Sachs, S. Weiler, T. Pietsch, U. Karsten, M. van de Wetering, H. Clevers, P. M. Schlag, W. Birchmeier, and J. Behrens.** 2002. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Mol. Cell. Biol. **22:**1184–1193.
32. **Polakis, P.** 2000. Wnt signaling and cancer. Genes Dev. **14:**1837–1851.
33. **Radtke, F., and H. Clevers.** 2005. Self-renewal and cancer of the gut: two sides of a coin. Science **307:**1904–1909.
34. **Reimand, J., M. Kull, H. Peterson, J. Hansen, and J. Vilo.** 2007. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. **35:**W193–W200.
35. **Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young.** 2000. Genome-wide location and function of DNA binding proteins. Science **290:**2306–2309.
36. **Sabates-Bellver, J., L. G. Van der Flier, M. de Palo, E. Cattaneo, C. Maake, H. Rehrauer, E. Laczko, M. A. Kurowski, J. M. Bujnicki, M. Menigatti, J. Luz, T. V. Ranalli, V. Gomes, A. Pastorelli, R. Faggiani, M. Anti, J. Jiricny, H. Clevers, and G. Marra.** 2007. Transcriptome profile of human colorectal adenomas. Mol. Cancer Res. **5:**1263–1275.
37. **Schones, D. E., A. D. Smith, and M. Q. Zhang.** 2007. Statistical significance of cis-regulatory modules. BMC Bioinform. **8:**19.
38. **Sierra, J., T. Yoshida, C. A. Joazeiro, and K. A. Jones.** 2006. The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes. Genes Dev. **20:**586–600.
39. **Spilianakis, C. G., and R. A. Flavell.** 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat. Immunol. **5:**1017–1027.
40. **Spilianakis, C. G., M. D. Lalioti, T. Town, G. R. Lee, and R. A. Flavell.** 2005. Interchromosomal associations between alternatively expressed loci. Nature **435:**637–645.
41. **Takahashi, M., Y. Nakamura, K. Obama, and Y. Furukawa.** 2005. Identification of SP5 as a downstream gene of the beta-catenin/Tcf pathway and its enhanced expression in human colon cancer. Int. J. Oncol **27:**1483–1487.
42. **Takemaru, K. I., and R. T. Moon.** 2000. The transcriptional coactivator CBP interacts with beta-catenin to activate gene expression. J. Cell Biol. **149:**249–254.
43. **Thorpe, C. J., G. Weidinger, and R. T. Moon.** 2005. Wnt/beta-catenin regulation of the Sp1-related transcription factor sp5l promotes tail development in zebrafish. Development **132:**1763–1772.
44. **van Beest, M., D. Dooijes, M. van De Wetering, S. Kjaerulff, A. Bonvin, O. Nielsen, and H. Clevers.** 2000. Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. J. Biol. Chem. **275:**27266–27273.
45. **Van der Flier, L. G., J. Sabates-Bellver, I. Oving, A. Haegebarth, M. De Palo, M. Anti, M. E. Van Gijn, S. Suijkerbuijk, M. Van de Wetering, G. Marra, and H. Clevers.** 2007. The intestinal Wnt/TCF signature. Gastroenterology **132:**628–632.
46. **van de Wetering, M., R. Cavallo, D. Dooijes, M. van Beest, J. van Es, J. Loureiro, A. Ypma, D. Hursh, T. Jones, A. Bejsovec, M. Peifer, M. Mortin, and H. Clevers.** 1997. Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. Cell **88:**789–799.
47. **van de Wetering, M., M. Oosterwegel, K. van Norren, and H. Clevers.** 1993. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. EMBO J. **12:**3847–3854.
48. **van de Wetering, M., I. Oving, V. Muncan, M. T. Pon Fong, H. Brantjes, D. van Leenen, F. C. Holstege, T. R. Brummelkamp, R. Agami, and H. Clevers.** 2003. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO Rep. **4:**609–615.
49. **van de Wetering, M., E. Sancho, C. Verweij, W. de Lau, I. Oving, A. Hurlstone, K. van der Horn, E. Batlle, D. Coudreuse, A. P. Haramis, M. Tjon-Pon-Fong, P. Moerer, M. van den Born, G. Soete, S. Pals, M. Eilers, R. Medema, and H. Clevers.** 2002. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell **111:**241–250.
50. **van Houte, L., A. van Oers, M. van de Wetering, D. Dooijes, R. Kaptein, and H. Clevers.** 1993. The sequence-specific high mobility group 1 box of TCF-1 adopts a predominantly alpha-helical conformation in solution. J. Biol. Chem. **268:**18083–18087.
51. **Wei, C. L., Q. Wu, V. B. Vega, K. P. Chiu, P. Ng, T. Zhang, A. Shahab, H. C. Yong, Y. Fu, Z. Weng, J. Liu, X. D. Zhao, J. L. Chew, Y. L. Lee, V. A. Kuznetsov, W. K. Sung, L. D. Miller, B. Lim, E. T. Liu, Q. Yu, H. H. Ng, and Y. Ruan.** 2006. A global map of p53 transcription-factor binding sites in the human genome. Cell **124:**207–219.
52. **Yang, A., Z. Zhu, P. Kapranov, F. McKeon, G. M. Church, T. R. Gingeras, and K. Struhl.** 2006. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. Mol. Cell **24:**593–602.
53. **Yochum, G. S., S. McWeeney, V. Rajaraman, R. Cleland, S. Peters, and R. H. Goodman.** 2007. Serial analysis of chromatin occupancy identifies beta-catenin target genes in colorectal carcinoma cells. Proc. Natl. Acad. Sci. USA **104:**3324–3329.
54. **Zheng, Y., S. Z. Josefowicz, A. Kas, T. T. Chu, M. A. Gavin, and A. Y. Rudensky.** 2007. Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. Nature **445:**936–940.

Chapter 5

Transcription factor Achaete scute-like 2 (Ascl2) controls intestinal stem cell fate

Cell, in press

Transcription factor Achaete scute-like 2 (Ascl2) controls intestinal stem cell fate

Laurens G. van der Flier[1], Marielle E. van Gijn[1], Pantelis Hatzis[1], Pekka Kujala[2], Andrea Haegebarth[1], Daniel E. Stange[1], Harry Begthel[1], Maaike van den Born[1], Victor Guryev[1], Irma Oving[1], Johan H. van Es[1], Nick Barker[1], Peter J. Peters[2], Marc van de Wetering[1] and Hans Clevers[1]

1) Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalaan 8, 3584 CT Utrecht, The Netherlands

2) The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Summary

The small intestinal epithelium is the most rapidly self-renewing tissue of mammals. Proliferative cells are confined to crypts, while differentiated cell types predominantly occupy the villi. We recently demonstrated the existence of a long-lived pool of cycling stem cells defined by *Lgr5* expression and intermingled with post-mitotic Paneth cells at crypt bottoms. We have now determined a gene signature for these so called Crypt Base Columnar (CBC) cells. One of the genes within this stem cell signature is the Wnt target *Ascl2*. Transgenic expression of the Ascl2 transcription factor throughout the intestinal epithelium induces crypt hyperplasia and *de novo* crypt formation on villi. Induced deletion of the *Ascl2* gene in adult small intestine leads to disappearance of the CBC stem cells within days. The combined results from these gain- and loss-of-function experiments imply that Ascl2 controls intestinal stem cell fate.

Introduction

The epithelium of the small intestine consists of crypt and villus domains. Crypts contain stem cells and their transit-amplifying (TA) daughter cells. Cells exiting the proliferative crypts onto the villi terminally differentiate into one of three specialised epithelial cell types: enterocytes, goblet cells and enteroendocrine cells. As an unusual cell type, Paneth cells escape the crypt-villus flow by migrating to crypt bottoms where they live for several weeks. With the exception of stem cells and Paneth cells, the intestinal epithelium is renewed approximately every 5 days (Barker et al., 2008). The primary driving force behind the proliferation of epithelial cells in the intestinal crypts is the Wnt signaling pathway. Mice that are mutant for an intestine-specific member of the Tcf transcription factor family, *Tcf4/Tcf7l2*, fail to establish proliferative crypts during late gestation (Korinek et al., 1998), while conditional deletion of ß-catenin (Ireland et al., 2004; Fevr et al., 2007) as well as transgenic expressing of the secreted Dickkopf-1 Wnt inhibitor (Pinto et al., 2003; Kuhnert et al., 2004) leads to disappearance of proliferative crypts in adult mice. Moreover, malignant transformation of intestinal epithelium is almost invariably initiated by activating Wnt pathway mutations (Korinek et al., 1997; Morin et al., 1997).

Because of the intimate connection between Wnt signaling and intestinal biology, we have attempted to unravel the Tcf4 target gene program activated by this pathway in crypts and colorectal tumors (van de Wetering et al., 2002; van Es et al., 2005a; Van der Flier et al., 2007). Through this approach, we have identified the *Lgr5* gene as a marker for CBC cells. Each crypt harbours around six of these small, cycling cells that are intermingled with Paneth cells at crypt bottoms. Using Cre-mediated genetic tracing, we demonstrated that CBC cells represent long-lived, multipotent stem cells (Barker et al., 2007), as predicted originally by Leblond and colleagues (Cheng and Leblond, 1974a; Cheng and Leblond, 1974b; Bjerknes and Cheng, 1981a; Bjerknes and Cheng, 1981b; Bjerknes and Cheng, 1999). Underscoring the "stemness" of $Lgr5^+$ CBC cells, we have recently establish an *in vitro* culture system in which a single sorted $Lgr5^+$ cell generates large organoids containing crypt and villus domains and harboring all differentiated cell types (Sato et al., Submitted). Although CBC cells occasionally occupy a position directly above the Paneth cells, they appear to be distinct from another proposed stem cell population specifically located at the so called +4 position (Potten et al., 1974; Potten, 1977), since CBC cells are not particularly radiation-sensitive and do not retain DNA labels (Barker et al., 2007; Barker et al., 2008). Bmi1 expression reportedly also marks cells at position +4. Lineage tracing has revealed that $Bmi1^+$ cells mark pluripotent stem cells that replenish the epithelium with similar kinetics to CBC cells (Sangiorgi and Capecchi, 2008). It remains to be resolved if Bmi1 and CBC cells are indeed distinct or may represent overlapping or even identical stem cell populations.

In a search for potential regulators of the intestinal stem cell fate, we have determined a gene expression profile for CBC cells, sorted on the basis of Lgr5-GFP expression from adult intestine. We paid particular attention to genes within this profile encoding transcription factors. The *Ascl2* gene was one of a few CBC cell-enriched genes that were not detected in the immediate daughters of these stem cells. Ascl2 (previously known as Mash2/HASH2) is one of the mammalian homologues of the *Drosophila* Achaete-scute complex genes (Johnson et al., 1990). *Ascl2* expression in the intestinal epithelium is Wnt-dependent (Sansom et al., 2004; Jubb et al., 2006; Van der Flier et al., 2007). The *Ascl2* gene encodes a basic helix-loop-helix (bHLH) transcription factor with an unusually restricted expression pattern, *i.e.* its expression is predominantly detected in extraembryonic tissues (Guillemot et al., 1994) and in intestinal epithelium (see below). Moreover, its expression is subject to paternal imprinting (Guillemot et al., 1995). Ascl2 plays an essential role in extraembryonic development, as evidenced by the observation that $Ascl2^{-/-}$ embryos die from placental failure around 10.5 days post-coitum (pc). In these mutant placentas, the spongiotrophoblasts lineage is depleted and the number of giant cells is elevated (Guillemot et al., 1994). Here we demonstrate an essential role for Ascl2 in the maintenance of adult intestinal stem cells through tissue-specific gain- and loss-of-function studies.

Results

Intestinal stem cell transcriptome

In order to define a gene expression profile for Lgr5^{+} intestinal stem cells, we established a protocol to sort GFP-positive epithelial cells from cell suspensions prepared from freshly isolated crypts of *Lgr5-EGFP-ires-CreERT2* mice (see Methods). As shown in Figure 1A, FACS analysis distinguished a GFP-high (GFP^{hi}) and a GFP-low (GFP^{lo}) population, which we tentatively identified as CBC cells and their immediate transit-amplifying daughters, respectively. A single mouse intestine routinely yielded several hundred thousand GFP^{hi} and GFP^{lo} cells. In order to identify novel stem cell genes, mRNA samples of the two populations were subjected to comparative gene expression profiling. A comprehensive list of CBC genes is given in Supplemental Table 1. The gene that was most highly enriched in the GFP^{hi} cells was, satisfactorily, the *Lgr5* gene itself. Multiple genes on the list were already identified as intestinal Wnt target genes previously (Van der Flier et al., 2007) which further validated the gene list (Suppl. Table 1). While *in situ* hybridizations on these Wnt target genes typically confirmed high level expression in CBC cells, TA cells directly above the Paneth cells often also expressed these genes, albeit at a lower level. As an example, Figure 1B shows the expression of *Sox9*, a Wnt-responsive gene (Blache et al., 2004) crucial for Paneth cell specification (Bastide et al., 2007; Mori-Akiyama et al., 2007).

Ascl2 was one of the known Wnt target genes (Sansom et al., 2004; Jubb et al., 2006; Van der Flier et al., 2007) and was expressed in adenomas as expected (Suppl. Figure 1A/D/E). Its physiological expression, however, was unique in that it was restricted to cells at the very base of crypts as revealed by *in situ* hybridization (Figure 1C). *In situ* hybridization analysis on a selected set of genes not observed previously as Wnt target genes (not shown) identified the *Olfactomedin-4* (*Olfm4*) gene as another highly-specific and robust marker for CBC cells (Figure 1D). Indeed, *Olfm4* appeared not to be expressed under the control of the Wnt pathway, since it was absent from adenomas (Suppl. Figure 1B). Human *OLFM4* was recently found to be enriched in microarray analysis of human colon crypt material (Kosinski et al., 2007). OLFM4 is a secreted molecule originally cloned from human myeloblasts (Zhang et al., 2002). No clear function has been assigned to this gene.

To study the expression pattern of *Ascl2* more precisely, we generated monoclonal antibodies against the C-terminal part of the mouse Ascl2 protein. One of these antibodies reacted specifically with spongiotrophoblasts cell nuclei in placental tissue of E10.5 in immunohistochemistry (IHC) (Suppl. Figure 1C). In the intestine, Ascl2 was found to be expressed in slender cells with the unique morphology and location of CBC cells, i.e. between the post-mitotic Paneth cells at the crypt base (Figure 1E), yet was not expressed in neighboring Paneth cells, or in the transit-amplifying cells which occupy the remainder of crypts. Figure 1F allows a comparison of Ascl2 nuclear staining to the cytoplasmic staining pattern for GFP in *Lgr5-EGFP-ires-CreERT2* knock-in intestine. Lgr5 and Ascl2 were expressed by the same CBC cells at the crypt base as revealed by serial sectioning (Suppl. Figure 1F/G).

To confirm the array results, we sorted GFP^{hi} and GFP^{lo} cells from duodenum, ileum and colon of *Lgr5-EGFP-ires-CreERT2* mice for real-time qPCR analysis. This confirmed that Lgr5, Ascl2, Tnfrsf19 and Olfm4 were highly enriched in stem cells of the small intestine (Suppl. Fig1H). Of note, only Lgr5 and Ascl2 were enriched in the GFP^{hi} stem cells of the colon. Expression of Bmi1, another putative stem cell marker, was readily detected in all fractions and was slightly enriched in GFP^{hi} stem cells of the small intestine, albeit to a lesser extent than Lgr5 or Ascl2. Of note, no such enrichment for Bmi1 was seen in the colon GFP^{hi} stem cells. We then tested intestines from 4 week-old *Bmi1* null animals (van der Lugt et al., 1994) for stem cell marker gene expression. Histologically and by marker analysis, the intestinal epithelium of these mutant mice was indistinguishable from wild type littermates (van Lohuizen and Clevers, unpublished). Both, Ascl2 and *Olfm4* were expressed in the intestines of these animals (Suppl. Figure 1I/J).

Consultation of EST data sets from man and mouse using the Unigene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene&cmd=search&term=) indicated that *Ascl2*

is expressed in only a very limited number of adult tissues, i.e. the small intestine and colon. Since our experiments indicated that the *Ascl2* gene resembled the *Lgr5* gene in its CBC cell-specific expression, we decided to investigate the function of *Ascl2* in the intestine in greater detail.

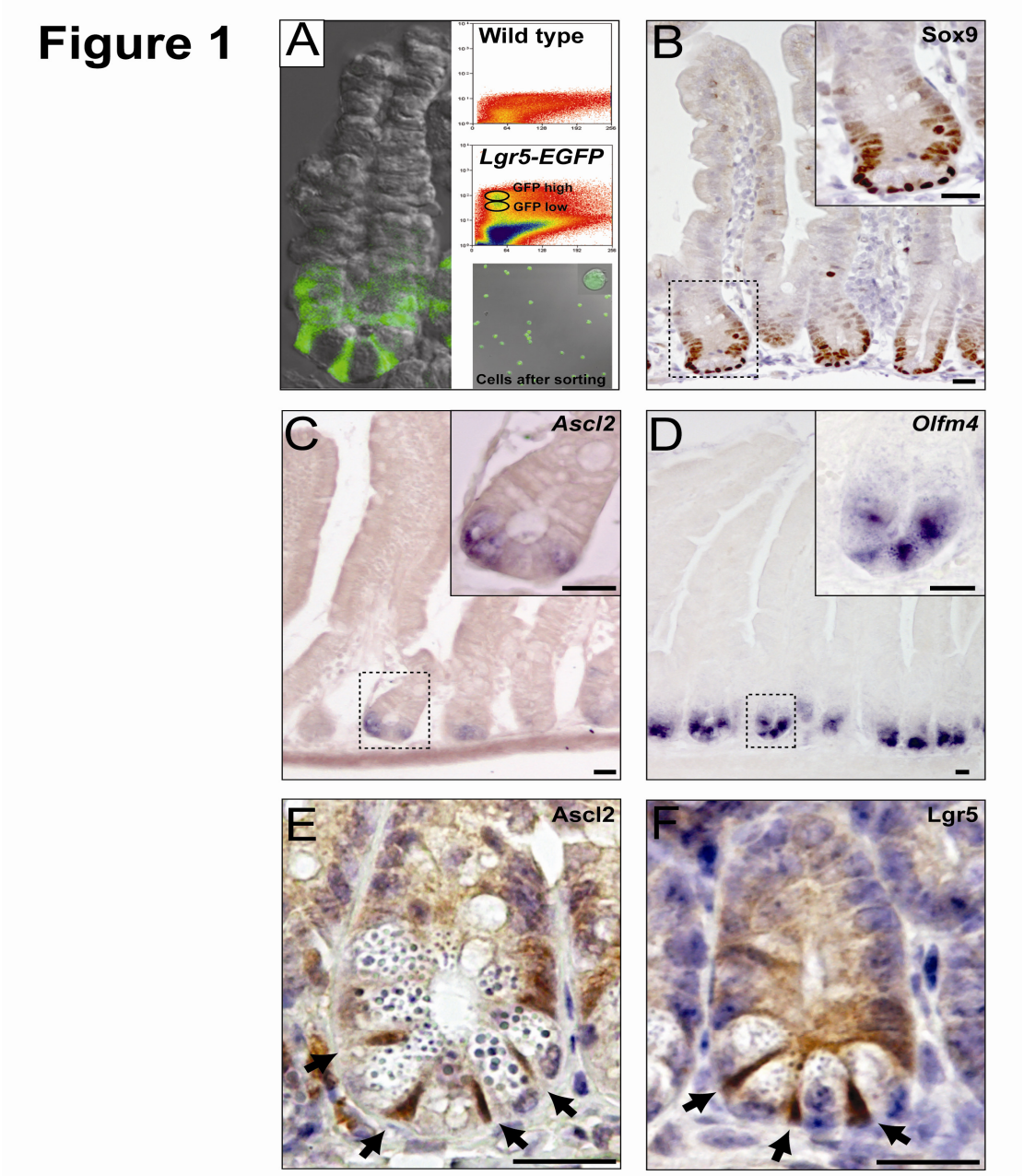


Figure 1 Ascl2, *Olfm4* and Lgr5 as intestinal stem cell markers
(A) Confocal images of an isolated *Lgr5-EGFP-ires-CreERT2* crypt (left). After treatment with trypsine, GFP^{hi} and GFP^{lo} cells were identifiable upon sorting (right/middle) when compared to wild type crypt cells (right/top). Confocal image of healthy sorted GFP^{hi} cells (right/bottom).
(B) Sox9 antibody staining shows high level expression in CBC cells and Paneth cells. TA cells directly above the Paneth cells also express the gene, albeit at a much lower level.
(C) In situ hybridization for *Ascl2* reveals a CBC cell-restricted expression pattern.
(D) In situ hybridization for *Olfm4* reveals a CBC cell-restricted expression pattern.
(E) Monoclonal Ascl2 antibody stains nuclei of CBC cells (arrows). Neither the surrounding post-mitotic Paneth cells at the crypt base, nor the transit-amplifying cells show any positive signal. Hematoxylin is used as a counter stain to visualize Paneth cell granules.
(F) Cytoplasmic GFP staining in *Lgr5-EGFP-ires-CreERT2* knock-in mice reveals CBC cells (arrows). Hematoxylin is used as a counter stain. Size bars represent 20 µm.

Transgenic expression of Ascl2

To initiate a functional study of Ascl2 in the intestine, we generated transgenic mice ectopically expressing a mouse *Ascl2* cDNA under the control of the villin promoter (insert in Figure 2B). This promoter drives expression throughout the intestinal epithelium and becomes fully active during late gestation (Pinto et al., 1999). Four independent transgenic mice were born. Three of these showed growth retardation and were sacrificed within 2-3 weeks postnatally. The founder of the fourth line was healthy and fertile, yet yielded multiple litters which contained transgenic F1 offspring at Mendelian ratios. These transgenic F1 individuals displayed the same phenotype as the three other independent transgenic mice. We concluded that the female surviving founder was either mosaic, or that –alternatively- the transgene in this particular founder mouse was variegated. In any case, this fortunate situation allowed us to study the phenotype in detail in the F1 offspring of the fourth transgenic line and to subsequently confirm the observations using fixed material from the three independent transgenic mice.

Transgenic animals were analyzed at P14 when they were still relatively healthy, yet had already replaced their postnatal intervillus pockets by adult-type crypts (Gregorieff and Clevers, 2005). To evaluate expression of the transgene, Ascl2 monoclonal antibody-staining was performed. As expected, nuclear Ascl2 was observed throughout crypt and villus epithelium in the transgenic mice (Figure 2B), while expression was restricted to crypt bottoms of the wild-type littermate controls (Figure 2A). H&E histology revealed dramatic changes in the intestinal epithelium of the Ascl2 transgenics. In the duodenum, villi were branched and displayed crypt-like pockets (Figure 2C/D). In the ileum, crypts were elongated ("hyperplastic"), while villi were short and disorganized compared to wild type controls (Figure 2E/F). Ki67 staining revealed that proliferation in the Ascl2 transgenic intestines was no longer restricted to the crypts, but also occurred in pockets along the villi (Figure 3 A/B). These regions displayed additional characteristics of crypts: the Wnt target gene *cMyc* for instance was also expressed in these proliferative pockets along the villi of the transgenic animals (Figure 3 E/F). *In situ* hybridizations for *Ets2*, a Wnt target gene specifically expressed at the crypt base (van de Wetering et al., 2002) and enriched in CBC cells (Suppl. Table 1) turned out to be also strongly expressed in the elongated crypts and along the villi of the transgenic animals (Figure 3 I/J). Of note, the villus pockets did not express the CBC-specific *Olfm4* and *Lgr5* genes. However, the expression domains of the latter two marker genes were expanded in the hyperplastic ileal crypts compared to control crypts as shown by *in situ* hybridizations in Figure 3 C, D, G, and H. A similar observation was made for the expression domain of the *Sox9* gene, which was dramatically expanded in the transgenic ileal crypts compared to wild type controls (Figure 3 K/L).

In the fly, the *achaete-scute* gene products form heterodimers with the Daugtherless protein (Cabrera and Alonso, 1991; Caudy et al., 1988). In a yeast 2-hybrid assay using ASCL2 as bait on a human colon library, we predominantly found the nuclear proteins E22, E2A and HEB as ASCL2-binding partners (van der Flier and Clevers, unpublished), confirming previous observations (Johnson et al., 1992; Scott et al., 2000). These three common E-proteins are the homologues of the fly gene product Daugtherless. In situ hybridization experiments on Apc^{min} intestines revealed expression of *E2a* and *Heb* in crypts and adenomas (Suppl. Fig 2A/B). The crypt-restricted expression of these co-factors explained why the transgenic overexpression of Ascl2 only partially affected the villus epithelium.

Figure 2

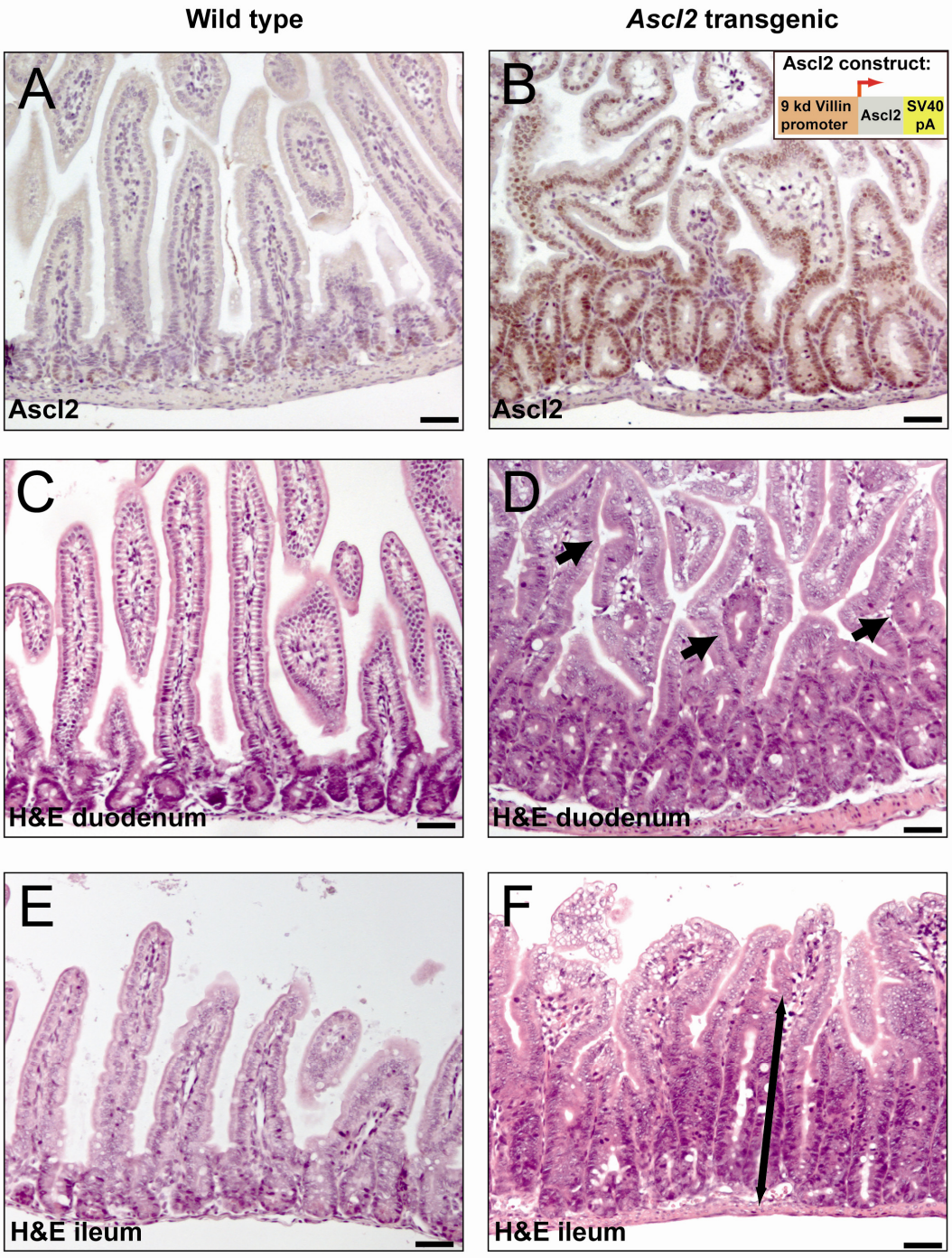


Figure 2 Intestine-specific Ascl2 miss-expression
(A) Monoclonal Ascl2 antibody staining reveals CBC-restricted Ascl2 expression in wild type animals.
(B) In *villin-Ascl2*-transgenic animals, epithelial nuclei along the crypt-villus axis stain for Ascl2. Insert outlines the transgene construct.
(C) H&E staining showing normal crypt-villus morphology in the duodenum of wild type animals.
(D) H&E staining showing aberrant epithelium morphology in the duodenum of transgenic animals. Villi are branched and display crypt-like pockets (arrows).
(E) H&E staining showing normal crypt-villus morphology in the ileum of wild type animals.
(F) H&E staining showing elongated crypts (arrow), with short, disorganized villi in the ileal epithelium of *Ascl2* transgenics. Size bars represent 50 μm.

Figure 3

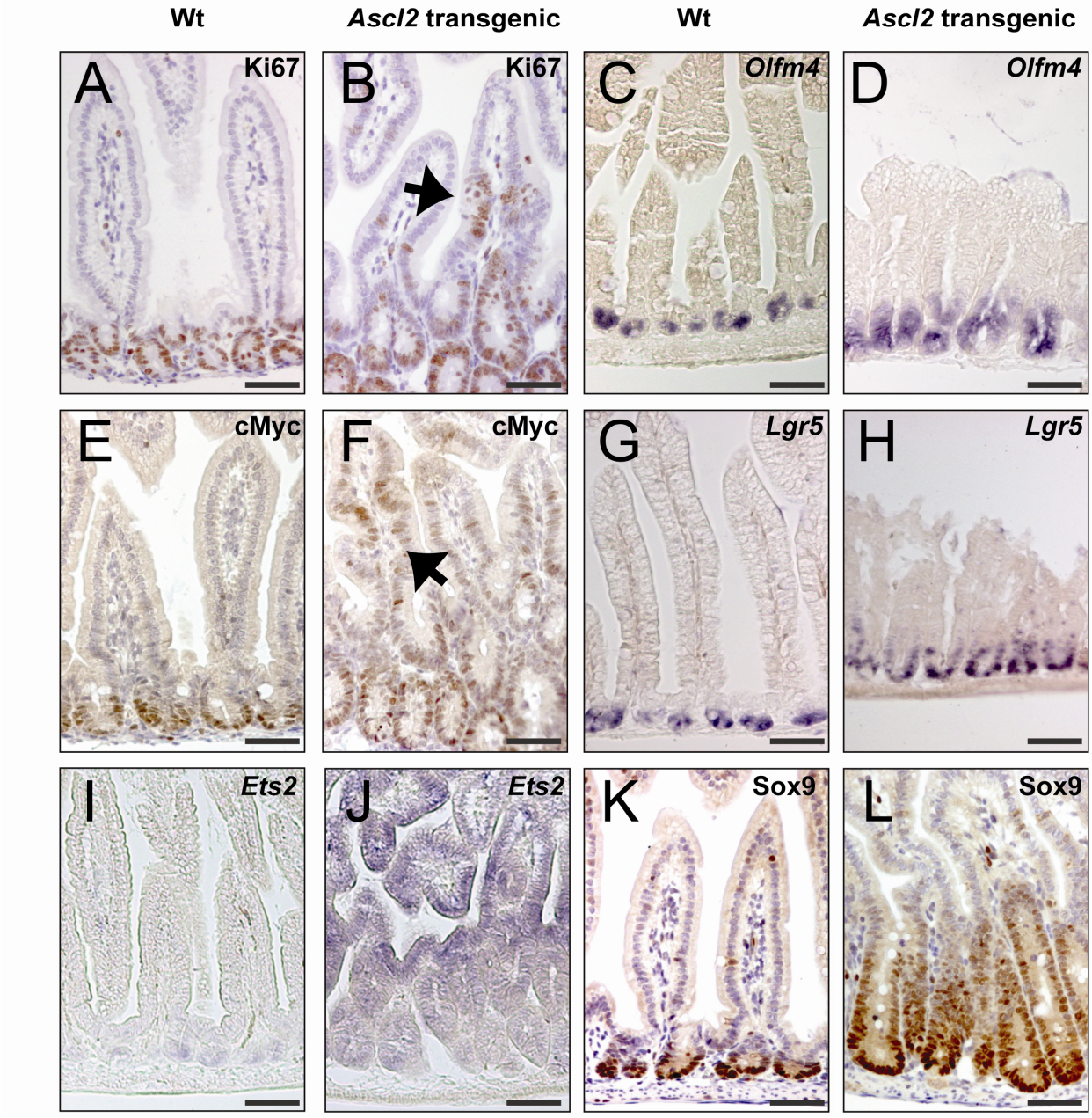


Figure 3 Marker analysis in the villin-Ascl2 intestine
Marker analysis of intestinal epithelium in wild type (A, C, E, G, I, K) and *Ascl2*-transgenic (B, D, F, H, J, L) animals.
(A, B) Ki67 analysis shows aberrant (arrow), dividing regions along the villus epithelium of *Ascl2* transgenic animals.
(C, D) *In situ* hybridization for the CBC marker *Olfm4* shows increased signal in the hyperplastic ileal crypts of *Ascl2* transgenics.
(E, F) The aberrant dividing regions along the villus epithelium of *Ascl2* transgenics are also positive for the crypt marker cMyc (arrow).
(G, H) *In situ* hybridization for CBC marker *Lgr5* shows elevated expression in the hyperplastic ileal crypts.
(I, J) *In situ* hybridization for the crypt marker *Ets2* show that the expression in *Ascl2* transgenic animals is found along the entire crypt-villus axis.
(K, L) IHC Staining for the crypt marker Sox9 shows expanded expression of this protein in the hyperplastic crypts of *Ascl2* transgenic animals. Size bars represent 50 µm.

Generation of a conditional *Ascl2* allele

Ascl2$^{-/-}$ animals die around 10.5 days *post coitum* due to a placenta defect (Guillemot et al., 1994). For this reason, we generated a conditional *Ascl2* allele, as depicted in Supplementary Figure 3A. The murine *Ascl2* gene contains three exons of which the middle exon encodes the conserved domains of the protein. The presence of an alternative promoter located between exon I and II has been suggested (Stepan et al., 2003). A conserved translation start occurs in exon 2, but the UCSC genome database annotates an alternative ATG start in exon 1 of the murine gene. For this reason, we introduced one LoxP site into the 5' UTR upstream of the alternative ATG in exon I. A second LoxP site was introduced between exon II and the untranslated third exon. Cre-mediated recombination should remove the entire open reading frame and result in a null allele. The paternal *Ascl2* allele is silenced during post-implantation development (Guillemot et al., 1995), a potential complication of our genetic analysis. By utilizing the heterozygous presence of the 5' UTR LoxP site as an allele-specific marker, we found that both alleles of *Ascl2* were expressed in the mouse intestine, implying that the paternal imprint was lost during development (Suppl. Figure 3B).

To study the role of Ascl2 in the intestine, we crossed the floxed animals with the *Ah-Cre* mouse (Ireland et al., 2004). In these mice, expression of the *Cre* recombinase is driven by the promoter of the *Cyp1a* gene, which is inducible in a number of tissues upon injection of lipophilic xenobiotic agents such as ß-Naphtoflavone (ßNF). In previous studies, it has been observed that when the *Ah-Cre* transgene is used to excise a roadblock sequence before LacZ and thus activate the Rosa26-LacZ reporter (Soriano, 1999), the resulting gene deletion in the intestinal epithelium is highly efficient and includes the intestinal stem cells (Ireland et al., 2004). The recombination remains stable over many months, implying that no selection pressure is exerted on the stem cells of these recombined LacZ+ crypts. A dramatically different observation was made when we deleted the functionally important *c-Myc* gene using the same protocol (Muncan et al., 2006). The epithelium, which was essentially c-Myc$^{-/-}$ at day 4 post-induction (PI), was entirely replaced by wild-type epithelium derived from low numbers of escaping, non-deleted cells within 2-3 weeks.

Initial experiments indicated that near-complete deletion of the *Ascl2*$^{floxed/floxed}$ allele occurred in the first few days PI. However, at late time points (>4 weeks), all signs of gene deletion had vanished (not shown). To narrow down this time-dependent effect, we analyzed adult *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals at 5, 8, 11 and 15 days PI and compared these to ßNF-treated *Ascl2*$^{floxed/floxed}$ litter mates. Intestinal epithelium was freed of associated stromal elements and was subjected to Southern blot analysis. The Southern strategy distinguished between wild-type, floxed, and floxed/recombined alleles as depicted in Supplementary Figure 3A. Figure 4A (upper panel) shows recombination of the *Ascl2*$^{floxed/floxed}$ locus in intestinal epithelial cells as observed at the indicated days PI. An apparently almost complete recombination was observed at the 5- and 8-day time points. At 11 days PI, a partial return of non-recombined alleles was observed. At day 15, recombined alleles were no longer observed. The *Ah-Cre* transgene is also inducible in the liver, albeit to a lesser extent (Ireland et al., 2004). As the *Ascl2* gene is not expressed in liver, its deletion should be neutral in terms of selection. As a control, we therefore analyzed *Ascl2* gene recombination in liver samples of the same animals. A significant, albeit not entirely complete, deletion of the *Ascl2* gene was observed (Figure 4A, lower panel). This gene deletion pattern in the liver remained unchanged over time, underscoring the high selective pressure observed in the intestine. The rapid reappearance of wild type epithelium implied a strong selective pressure favoring the few remaining Ascl2^{+} epithelial CBC cells. Quantitative recombination was confirmed through Northern blot analysis of intestinal epithelial extracts (Figure 4B). Ascl2 monoclonal antibody stainings (Figure 4C-G) on intestines from induced animals also confirmed the absence and reappearance of the Ascl2 protein in CBC cells in time post-induction.

To directly visualize the rescue process, we bred the Cre-activatable Rosa26-LacZ reporter (Soriano, 1999) into the *Ah-Cre/Ascl2*$^{floxed/floxed}$ strain. *Rosa26-LacZ*|*Ah-Cre/Ascl2*$^{floxed/floxed}$ animals were treated with ßNF and intestines were isolated at 5, 10 or 20 days PI. Intestines of *Rosa26-LacZ*|*Ah-Cre/Ascl2*$^{floxed/wt}$ animals were used as controls. Figure 4H-K illustrates LacZ staining of the duodenum. At 5 days PI, virtually all cells were recombined and blue except for long-lived Paneth cells (Figure 4H). At 10 days PI, non-recombined epithelial cells started reappearing (Figure 4I). At 20 days PI, virtually all crypts were again wild type (Figure 4J). In contrast, in the

liver of this animal, blue cells still predominated (data not shown). In heterozygous intestines at 20 days PI, the entire epithelium stained blue (Figure 4K).

Figure 4

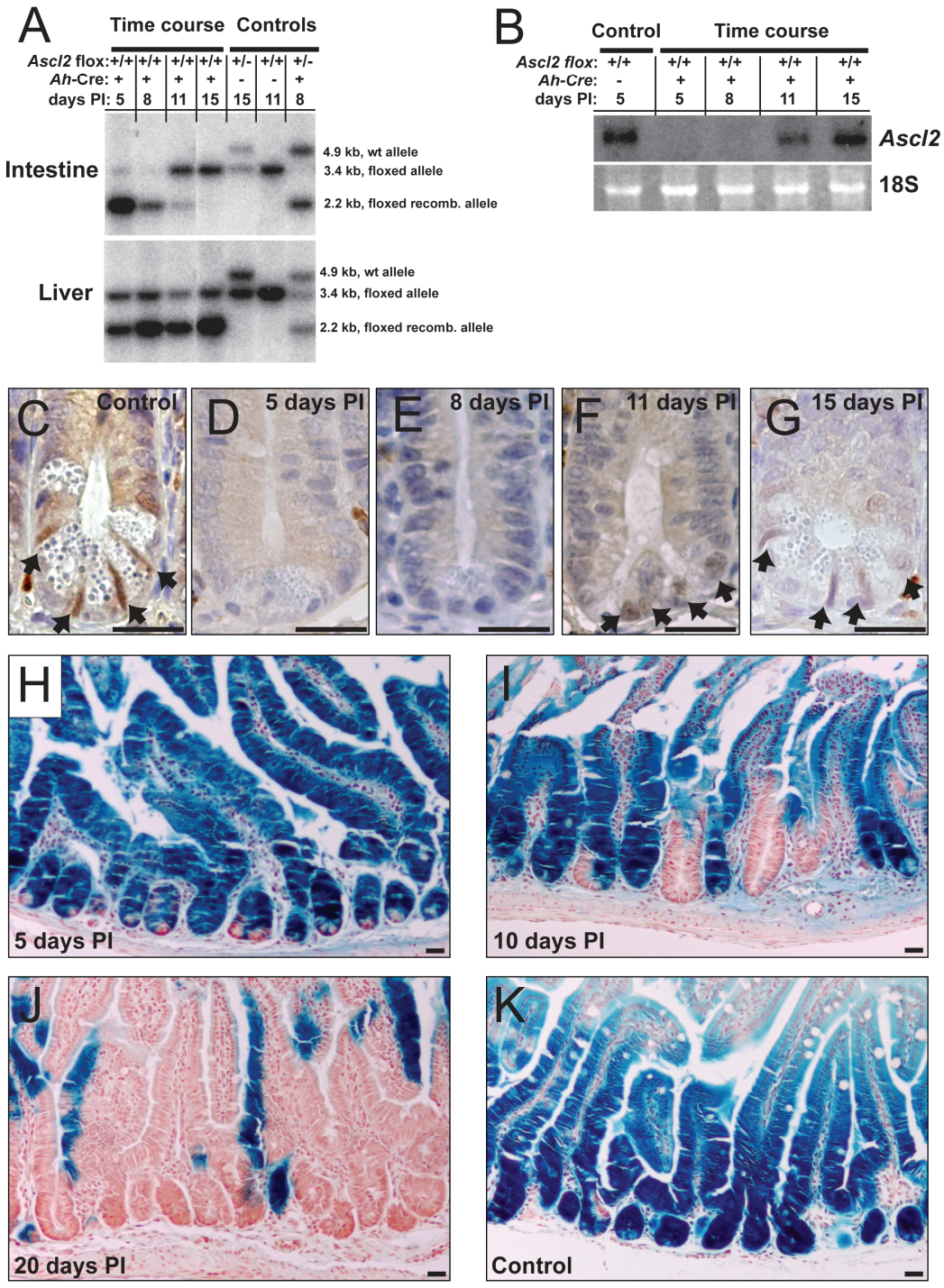


Figure 4 Conditional *Ascl2* inactivation in the intestinal epithelium

(A) Southern blot analysis of genomic DNA of intestinal epithelial cell extracts (upper panel) and liver samples (lower panel) of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 5, 8, 11 and 15 days PI and control animals. DNA was digested with EcoRV and hybridized with probe A (as indicated in Supplementary Figure 3A).

(B) *Ascl2* Northern blot analysis on total RNA from intestinal epithelial cell extracts of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 5, 8, 11 and 15 days PI and control.

(C-G) Monoclonal Ascl2 antibody staining demonstrates CBC restricted expression of the Ascl2 protein (arrows) in control animals (C). 5 (D) and 8 (E) days PI the Ascl2 signal is completely gone. The signal reappears in the CBC cells 11 (F) and 15 (G) days PI.

(H-K) LacZ staining of the duodenum illustrates recombination efficiency in *Rosa26-LacZ/Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 5, 10 and 20 days PI. At 5 days PI, virtually all cells were recombined (blue) except for long-lived Paneth cells (H). At 10 days PI, non-recombined epithelial cells started reappearing (I). At 20 days PI, virtually all crypts were wild type (J). In *Rosa26-LacZ/Ah-Cre/Ascl2*$^{floxed/wt}$ control animals 20 days PI, the entire intestinal epithelium remained blue (K). Size bars represent 25 μm.

Figure 5

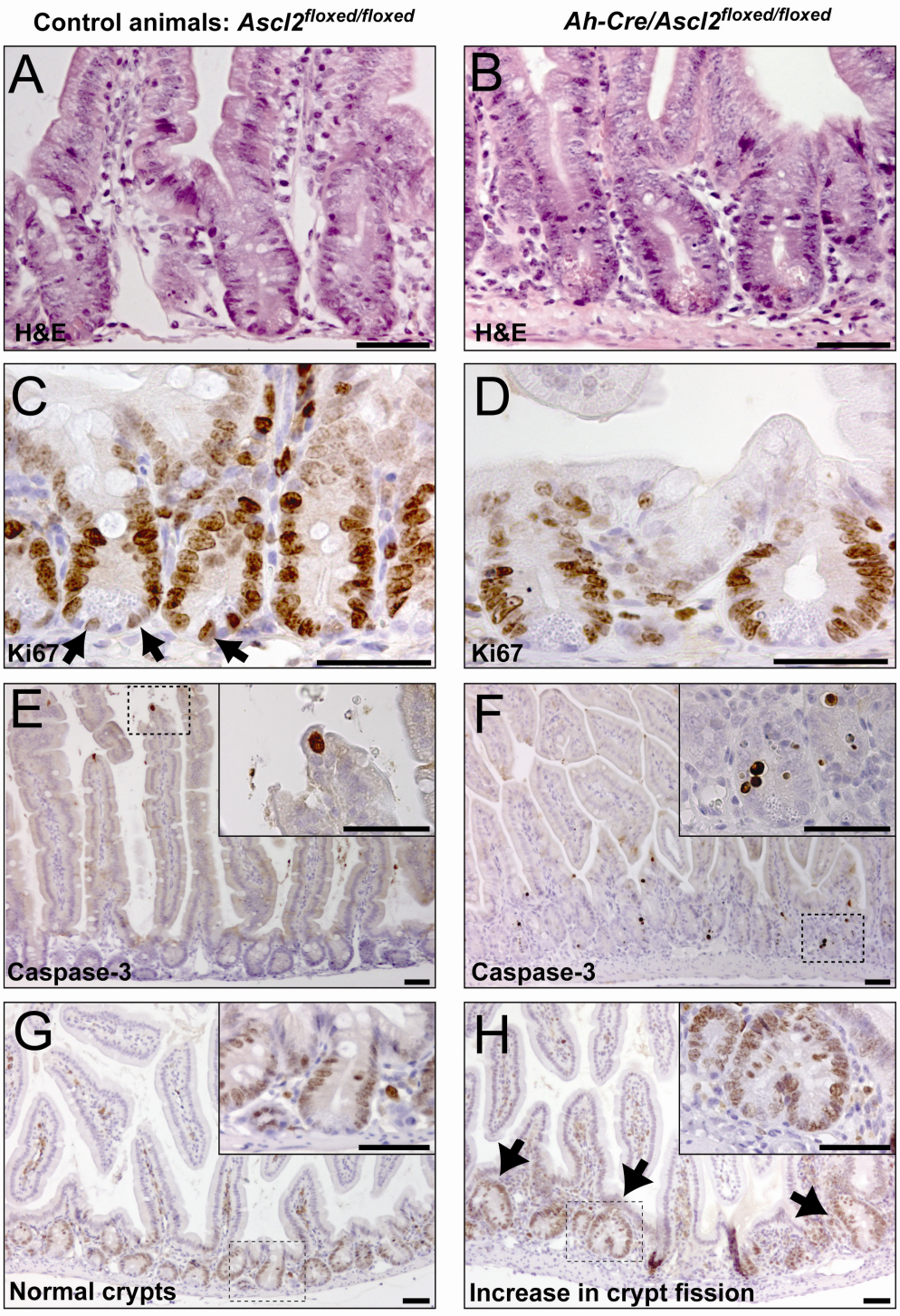


Figure 5 Expansion of Ascl2 wild-type crypts through crypt fission

(A, B) H&E staining reveals normal morphology of crypt-villus epithelium of *Ascl2*$^{floxed/floxed}$ control animals (A) and *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 5 days PI (B).

(C, D) Ki67 staining showing normal numbers of proliferating cells in intestinal crypts of *Ascl2*$^{floxed/floxed}$ control animals (C) and *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 5 days PI (D).

(E, F) Active Caspase-3 staining showing apoptosis predominantly at the villus tip in *Ascl2*$^{floxed/floxed}$ control animals (E). At 5 days PI, a strong increase in apoptosis in the crypt region of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals can be observed (F).

(G, H) Ki67 staining showing at a low magnification, normal crypt-villus epithelium in *Ascl2*$^{floxed/floxed}$ control animals (G). An increase in crypt fission (arrows) in *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals 11 days PI is observed (H). Size bars represent 50 μm.

Intestinal stem cells are lost upon conditional deletion of *Ascl2*

No dramatic histological differences were observed between the *Ah-Cre/Ascl2*$^{floxed/floxed}$ and control animals by H&E staining (Figure 5A/B) or Ki67 staining (Figure 5C/D) at 5 days PI. Mutant crypts were generally broader at their base (Figure 5C/D). The crypt base entirely consisted of Paneth cells, while CBC cells recognizable by their elongated Ki67^{+} nuclei were conspicuously absent (Figure 5C/D). These observations were confirmed by electron microscopy (Suppl. Figure 4A/B). There was an increase in apopototic cells in the crypts of floxed animals as revealed by active Caspase-3 staining 5, 8 and 11 days PI (Figure 5F). Normally, apoptotic cells in the intestine are predominantly observed at the villus tips (Figure 5E). Analysis at late time points PI revealed a strong increase in crypt fission profiles (Figure 5 G/H). Crypt fission represents a powerful repair mechanism of the intestinal epithelium, activated for instance after radiation injury or genetic damage (Cairnie and Millen, 1975; Muncan et al., 2006).

To follow the fate of CBC cells more precisely, we utilized the robust CBC marker *Olfm4* (see Figure 1D). *In situ* hybridizations for *Olfm4* expression confirmed that the *Olfm4*$^{+}$ CBC cells were present in every single crypt (See Figure 6A for a low power magnification). Five days PI, *Olfm4*$^{+}$ CBC cells had largely vanished (Figure 6B). Of note, very rare *Olfm4*$^{+}$ cells remained (black arrow in Figure 6B). At day 8 PI, the large majority of crypts were still completely negative, yet an Olfm4 signal reappeared in occasional crypts. Some crypts only contained 1 or 2 positive CBC cells (black arrow in Figure 6C), while others presented normal wild type *Olfm4* staining profiles (Figure 6C). By 11 days PI (Figure 6D), larger patches of *Olfm4*$^{+}$ crypts reappeared, underscoring the activation of crypt fission-based repopulation. By day 15 PI, all crypts were *Olfm4*$^{+}$ (Figure 6E). Thus, the presence of *Olfm4*$^{+}$ CBC cells closely paralleled the presence of intact *Ascl2* alleles.

To investigate the target gene program regulated by Ascl2 in the small intestine, we performed comparative gene expression profiling on RNA samples from isolated intestinal epithelium of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals and *Ah-Cre/Ascl2*$^{floxed/wt}$ control animals at day 3 and 5 days PI. A total of 130 genes, also expressed by CBC cells, was significantly downregulated > 1.5-fold at both time points PI. Comparison of these Ascl2 target genes with the stem cell gene signature (Supplemental Table 1) results in a ~25% overlap (Figure 7A). The identity of the genes within this overlap is given in Supplemental Table 2.

To identify in vivo ASCL2 binding sites, we performed chromatin immunoprecipitations (ChIP) on human LS174T colorectal cancer (CRC) cells using one of the monoclonal Ascl2 antibodies. In a pilot experiment, we probed a customized array containing a set of TCF4 regulated genes including many of the stem cell-enriched genes given in Supplemental Table 2 (unpublished results Hatzis, Van der Flier and Clevers). Potential ASCL2-bound regulatory regions were verified by qPCR. LGR5, EPHB3 and TNFRSF19 promoters were bound by ASCL2, while the OLFM4 promoter was not. Non-promoter sequences in the PTPRO, SOAT1, ETS2 and SOX9 genes were also bound by ASCL2 (Figure 7B). As a further validation of these data, we created an LS174T transfectant carrying a stably integrated, doxycycline-inducible sh-RNA expression vector targeting ASCL2. An essentially complete knock-down of the ASCL2 mRNA occurred within 24 hours upon RNAi induction, as shown by Northern blot analysis (Suppl. Figure 5A). We used this cell line to validate the generated ChIP signals. As can be seen in Supplementary Figure 5B, the ChIP signals for the EPHB3 promoter and the ETS2 enhancer almost entirely disappear upon ASCL2 knock-down.

As one example of the functional activity of an ASCL2-bound genomic region, we cloned the putative SOX9 enhancer as a 1 kb fragment in a luciferase reporter plasmid. Transient transfection was performed in the inducible LS174T ASCL2 RNAi cell line. Figure 7C shows that the spontaneous activity of the SOX9 enhancer was significantly downregulated upon ASCL2 knock-down. For the reverse experiment we stably integrated a dox-inducible ASCL2 expression plasmid into HCT116 cells. This CRC cell line does not express endogenous ASCL2 (Suppl. Figure 5C). Transient transfection of the SOX9 enhancer plasmid into this cell line yielded in a significant activation of the SOX9 enhancer upon ASCL2 induction (Figure 7D).

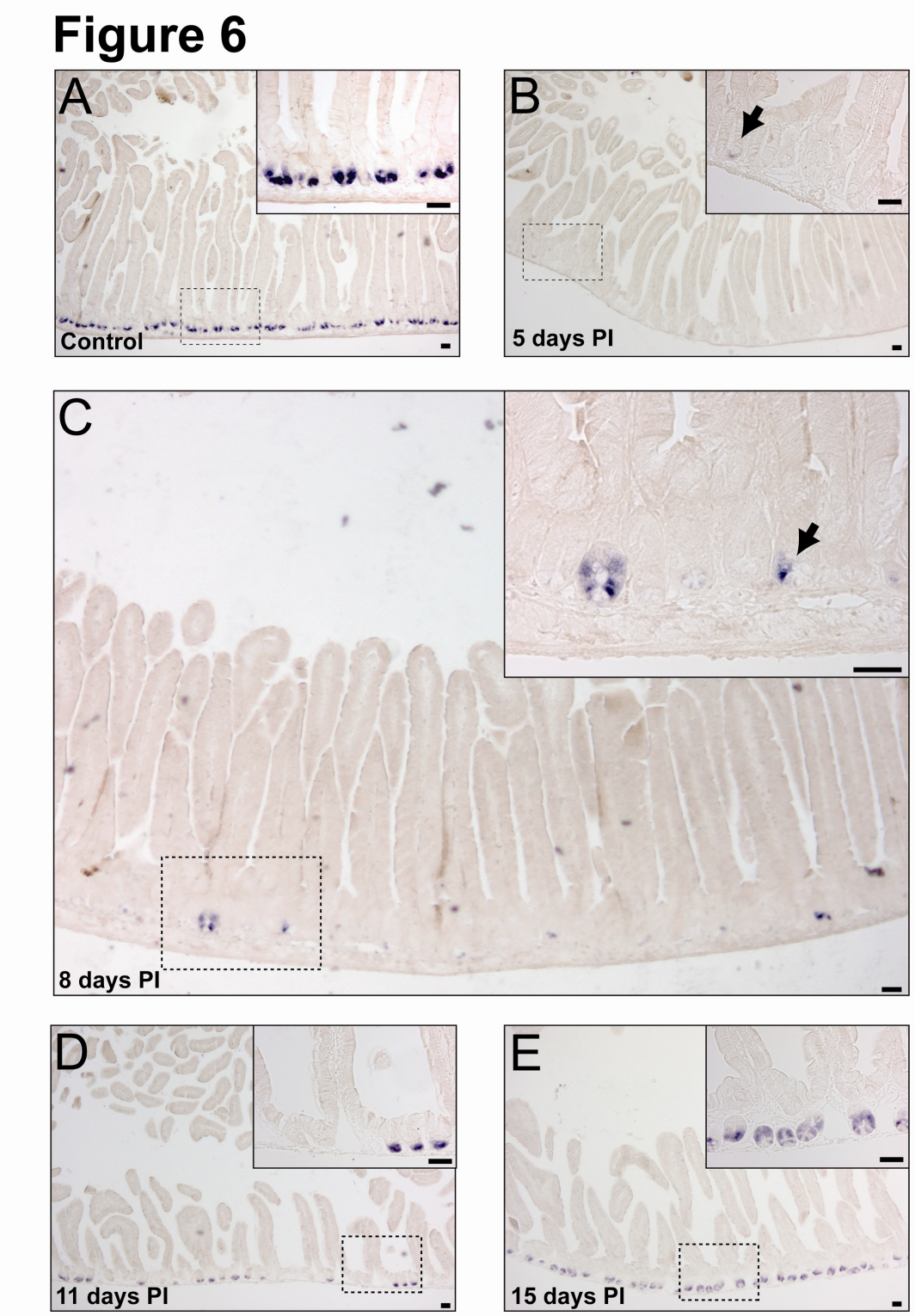


Figure 6 Intestinal stem cells are lost upon Ascl2 deletion
Olfm4 in situ hybridizations to visualize CBC cells in *Ah-Cre/Ascl2*$^{floxed/floxed}$ small intestines 5, 8, 11 and 15 days PI and control intestine.
(A) *Olfm4* in situ staining at low magnification shows CBC specific staining in *Ascl2*$^{floxed/floxed}$ control animals in every crypt.
(B) The *Olfm4* signal is almost completely gone 5 days PI in *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals. Sporadic positive CBC cells can be observed (arrow).
(C) Eight days PI most of the crypts are still negative for *Olfm4*, although sporadic crypts display normal staining, while other crypts contain 1 or 2 positive CBC cells (arrow).
(D) At 11 days PI, the tissue of *Ah-Cre/Ascl2*$^{floxed/floxed}$ shows patches of positive and patches of negative crypts for *Olfm4*.
(E) The *Olfm4* signal is completely restored 15 days PI in the *Ah-Cre/Ascl2*$^{floxed/floxed}$ *animals*. Size bars represent 50 μm.

We interpreted the combined observations in the following manner. The *Ascl2* gene is crucial for the maintenance of CBC cells, the only cells in the intestine that express the gene. Part of the stem cell signature is Ascl2 dependent and involves binding of Ascl2 to promoters and/or enhancers elements to activate transcription of these target genes. In the absence of Ascl2 expression, CBC cells rapidly disappear, leading to a reshaping of the crypt bottom which becomes filled with long-lived Paneth cells only. *Ascl2* deletion in TA cells does not directly harm these cells as they normally don't express the gene. The mutant TA cells continue to proliferate and provide the villi with differentiated cells. However, no new TA cells are produced from CBC cells in mutant crypts. The strong increase in apoptosis in the TA compartment likely mirrors the limited self-renewal capacity that these cells possess. As a consequence, a strong selective pressure is exerted which favors the re-emergence of "escaper" crypts harboring intact *Ascl2* alleles.

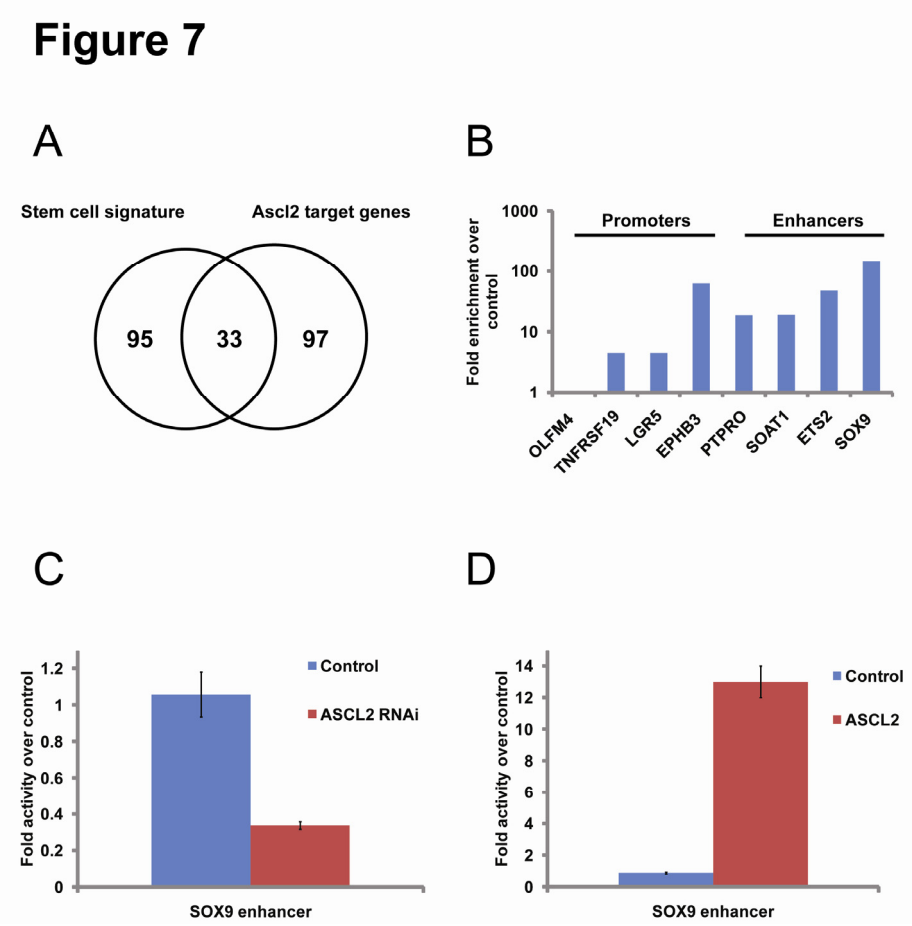


Figure 7 Ascl2 target genes

(A) Venn diagram showing the overlap between the stem cell signature (Supplemental Table 1) and the Ascl2 target genes. Ascl2 target genes have been identified through comparative gene profiling of RNA samples from isolated intestinal epithelium of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals and *Ah-Cre/Ascl2*$^{floxed/wt}$ control animals at day 3 and 5 days PI. For the comparison of the stem cell genes and the Ascl2 target genes only those genes expressed in both lists were compared. The identity of the overlapping genes is given in Supplemental Table 2.

(B) Association of ASCL2 with the proximal promoters and enhancers of the indicated Ascl2 target genes. ASCL2 binding to indicated regions is expressed as relative enrichment of the respective qPCR product over the qPCR product of the non-bound exon2 of the myoglobin gene. Experiments have been performed at least three times. A representative experiment is shown.

(C) Transcriptional activity of SOX9 enhancer in integrated inducible ASCL2 RNAi LS174T cell line. Luciferase reporter containing the SOX9 enhancers was transiently transfected in the LS174T ASCL2 RNAi cell line. A CMV-Renilla reporter was cotransfected as a normalizing control. Cells were induced for 24 hr with doxycycline to induce the sh-RNA against ASCL2. Control and ASCL2 RNAi values are normalized over the empty pGL4.10 TATA construct. Experiments have been performed at least three times. A representative experiment is shown. Values are the average of normalized triplicates and error bars represent standard deviations of these triplicates.

(D) Transcriptional activity of SOX9 enhancer in stably integrated inducible ASCL2 HCT116 cell line. Luciferase reporter containing the SOX9 enhancer was transiently transfected in inducible ASCL2 HCT116 CRC cell line. A CMV-Renialla reporter was cotransfected as a normalizing control. Cells were induced for 24 hr with doxycycline to induce ASCL2 expression. Control and ASCL2 induction values are normalized over the empty pGL4.10 TATA construct. Experiments have been performed at least three times. A representative experiment is shown. Values are the average of normalized triplicates and error bars represent standard deviations of these triplicates.

Discussion

In the intestine, a long-lived pool of cycling stem cells is defined by *Lgr5* expression, a Wnt responsive orphan G-coupled receptor (Barker et al., 2007). Here, we define a minimal gene expression profile for these CBC cells by exploiting the *Lgr5-EGFP-ires-CreERT2* knock-in mice for sorting, based on GFP expression. We describe a set of about 130 genes differentially expressed between GFP^{hi} and GFP^{lo} fractions. Based on confocal images of isolated crypts, we tentatively identified these as CBC cells and their daughters, respectively. Lgr5 is the most differentially expressed gene within this set. Many other genes in the signature represent previously identified Wnt-dependent genes, e.g. *Ascl2, CD44, Ephb3* and *Sox9* (Van der Flier et al., 2007). Given the intimate connection between Wnt signaling and the biology of stem cells in many tissues (Reya and Clevers, 2005), this was not surprising. At least one novel marker for CBC cells, *Olfm4*, was not expressed under the control of the Wnt pathway, implying the existence of a Wnt-independent specifier of CBC identity. We do not currently know the molecular nature of this specifier, but assume that it may be driven by signal(s) from the stem cell niche.

In order to detail the molecular identity of CBC stem cells, we focused on genes within the signature which displayed a CBC-restricted (vs. enriched) expression. Monoclonal antibody staining for Ascl2 confirmed that this protein is only expressed by the CBC cells of the intestine. Ascl2 is one of the mammalian homologous of the *Drosophila achaete-scute* complex genes (Johnson et al., 1990). This complex consists of four physically linked genes that encode related bHLH proteins and represent powerful regulators of cell fate. The *achaete-scute* genes are initially expressed in proneural cell clusters were they promote neuroblast differentiation and are essential for the differentiation of the central as well as peripheral nervous system (Calleja et al., 2002). *Achaete-scute* genes are best known as targets of the Notch pathway. *Hairy/Enhancer of Split* genes, activated by Notch signaling, directly repress the proneural *achaete-scute* genes. The indirect repression by Notch signals allows *achaete-scute* gene-expressing cells to be singled-out through lateral inhibition (Simpson, 1990). However, *achaete-scute* genes can also be expressed under control of the Wnt pathway. Wingless expression is required for *achaete-scute* complex gene expression along the wing margin and for *achaete-scute*-dependent formation of margin bristles and their precursors (Phillips and Whittle, 1993). Notch signaling does not appear to regulate *Ascl2* expression in the mammalian intestine. Intestinal specific loss-of-function studies for Notch show a complete conversion of intestinal epithelium into goblet cells (van Es et al., 2005b; Riccio et al., 2008). If Ascl2 would be regulated by the Notch pathway, inhibition of Notch signaling would have resulted in upregulation of Ascl2, which we have not observed (van Es and Clevers, unpublished). Rather, *Ascl2* is expressed under the control of the Wnt pathway (Sansom et al., 2004; Jubb et al., 2006; Van der Flier et al., 2007). In a recent genome-wide TCF4 Chromatin Immunoprecipitation study, we have observed that the *ASCL2* locus in human CRC cells contains two TCF4/ß-catenin-bound regulatory elements, indicating that ASCL2 is a direct target of the Wnt pathway (Hatzis et al., 2008).

Wnt target genes in the intestine come in several expression categories (Van der Flier et al., 2007). While most are expressed in TA cells, a small set of transcriptional Wnt targets is expressed in the terminally differentiated Paneth cells, while others –like *Ascl2* and *Lgr5*- are expressed specifically in CBC cells. We also observe combinations of these expression domains: e.g. *EphB3* is expressed in CBC cells and Paneth cells, while *EphB2* is expressed in CBC cells and all TA cells (Batlle et al., 2002; and this study). While we do not know currently how individual Wnt target genes are restricted in their expression to different cell types within the intestinal crypts, this appears to be an important issue as it may explain fundamental differences between long-lived stem cells, short-lived TA cells and terminally differentiated Paneth cells.

Like its fly counterparts, mammalian Ascl2 appears to control lineage specification. In *Ascl2*-mutant mice, the spongiotrophoblast layer is lost by embryonic day E10 at the expense of trophoblast giant cells (Guillemot et al., 1994). Indeed, the main function of placental Ascl2 appears to rest in the maintenance of spongiotrophoblasts (Hughes et al., 2004). The current study demonstrates that *Ascl2* is essential for the maintenance of CBC stem cells in the adult intestinal epithelium and that miss-expression of *Ascl2* in non-CBC cells results in crypt hyperplasia and the presence of proliferative, crypt-like pockets on villi. We note that *Ascl2* expression alone does not

convert any intestinal epithelial cells into a *bona fide* CBC cell. Rather, a number of features of stem cell compartments are induced or enhanced by *de novo* expression of *Ascl2*.

In conclusion, we have initiated a molecular characterization of the stem cell of the intestinal epithelium by gene expression profiling of essentially pure CBC cells. The current study of Ascl2 demonstrates that, with the available genetic tools, the intestinal epithelium constitutes a prime model to unveil molecular mechanisms underlying the biology of stem cells.

Experimental Procedures

Isolation of GFP positive epithelial cells
Freshly isolated small intestines were incised along their length and villi were removed by scraping. The tissue was then incubated in PBS/EDTA (5 mM) for 5 minutes. Gentle shaking removed remaining villi and the intestinal tissue was subsequently incubated in PBS/EDTA for 30 minutes at 4 ^{o}C. Vigorous shaking, yielded free crypts which were incubated in PBS supplemented with Trypsine (10 mg/ml) and DNAse (0.8 u/µl) for 30 minutes at 37 ^{o}C. After incubation, cells were spun down, resuspended in SMEM (Invitrogen) and filtered through a 40 µM mesh. GFP-expressing cells were isolated using a MoFlo cell sorter (DAKO).

Microarray analysis
For the stem cell signature RNA was isolated from sorted GFP^{hi} and GFP^{lo} cell fractions of intestines from *Lgr5-EGFP-ires-CreERT2* mice. For the analysis of Ascl2 target genes RNA was isolated from intestinal epithelial cells of *Ah-Cre/Ascl2$^{floxed/floxed}$* animals and *Ah-Cre/Ascl2$^{floxed/wt}$* control animals 3 and 5 days PI. 500 ng of total RNA was labeled using low RNA Input Linear Amp kit (Agilent Technologies, Pato Alto, CA, USA). Labeling, hybridization, and washing protocols were done according to Agilent guidelines. Differentially labelled cRNA from GFP^{hi} and GFP^{lo} cells from two different sorts (each combining three different mice) were hybridised on 4X44K Agilent Whole Mouse Genome dual colour Microarrays (G4122F) in two dye swap experiments, resulting in four individual arrays. For the Ascl2 target gene analysis we analyzed the 3 and 5 days PI experiments in two dye swap experiments, resulting in four individual arrays. Microarray signal and background information were retrieved using featuere extraction (V.9.5.3, Agilent Technologies). All data analyses were performed using ArrayAssist (5.5.1, Stratagene Inc, La Jolla, CA, USA) and Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). Raw signal intensities were corrected by subtracting local background. Negative values were changed into a positive value close to zero (standard deviation of the local background) in order to allow calculation of ratios between intensities for features only present in one sample (GFP^{hi} or GFP^{lo}) or (*Ah-Cre/Ascl2$^{floxed/wt}$* or *Ah-Cre/Ascl2$^{floxed/floxed}$*). Data were filtered if both (GFP^{hi} and GFP^{lo}) or (*Ah-Cre/Ascl2$^{floxed/wt}$* or *Ah-Cre/Ascl2$^{floxed/floxed}$*) intensities were changed or if both intensities were less than two times the background signal and normalized using the Loess algorithm. Array data will be available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) upon publication. Statistical analysis was performed with SAM (Significant Analysis of Microarrays) using an Excel plug-in of the software (Tusher et al., 2001) and "one class" as the response value. For the stem cell signature genes were considered to be significantly enriched in GFP^{hi} cells if they were present (not filtered) in all four arrays, had a q-value of <10, where enriched (>log2 ratio of 0.58) in at least 2 out of 4 arrays and had an average of all four arrays exceeding a log2 ratio of 0.58. Genes were considered to be Ascl2 targets when significantly down regulated in *Ah-Cre/Ascl2$^{floxed/floxed}$* intestines, if they were present (not filtered) in all four arrays, had a q-value of <7.5, where down regulated (<log2 ration of -0.58) in at least 2 out of 4 arrays and had an average of all four arrays below a log2 ratio of -0.58. For the comparison of the Ascl2 target genes and stem cell genes only those genes present (not filtered) in both lists were compared.

cDNA synthesis
cDNA was synthesized from 0.3-1 µg of total RNA of GFP^{hi} and GFP^{lo} cells derived from duodenum, ileum and colon of *Lgr5-EGFP-ires-CreERT2* mice, using 0.5 µg random primers (Promega) per synthesis according to manufacturer's instructions.

Generation of monoclonal Ascl2 antibody
A PCR product encoding an 87 amino acid C-terminal part of the mouse Ascl2 protein was cloned into pET21b prokaryote expression vector (Novagen) and used to generate His-tagged recombinant protein in BL21 E. coli. Six-week-old BALB/c mice were immunized with 50 µg of purified C-terminal Ascl2 fusion protein in Freund's incomplete adjuvant (Difco) three times. Following sacrifice of immunized mice, the splenocytes were fused with SP2/0 mouse myeloma cells using polyethylene glycol. The fused cell population was cultured in hypoxanthine aminotropterin thymidine selection medium (Invitrogen) and plated

into 96-well flat bottom culture plates. Hybridoma supernatants were screened by immunochemical staining of methanol-fixed COS cells transfected with a mouse Ascl2 expression plasmid. Two positive clones were found, of which one also recognizes the human ASCL2 protein.

Generation of Villin-Ascl2 mice
The *villin-Ascl2* transgenic expression construct was generated by cloning the coding mouse Ascl2 sequence at the initiation codon of the 9-kb regulatory region of the mouse villin gene (Pinto et al., 1999) using standard techniques. An SV40 termination and polyadenylation cassette was added downstream (Insert in Figure 2B). *Villin-Ascl2* transgenic mice were generated by microinjection of linearized plasmid into the pronuclei of fertilized eggs of B6CBAF1/JIco mice. Transgenic mice were identified by Southern blotting and PCR analysis using tail genomic DNA. Founder mice were bred to C57BL/6 mice.

Generation of *Ascl2*$^{floxed/floxed}$ mice
The conditional *Ascl2* targeting construct was generated by cloning the various components into the pl451 vector as depicted in Supplementary Figure 3A. Both flanking arms and the exon of Ascl2 were generated by high-fidelity PCR reactions from male 129/Ola-derived IB10 embryonic stem cell DNA. All components were sequence-verified. The targeting construct (100 µg) was linearized and transfected into male 129/Ola-derived IB10 embryonic stem cells by electroporation (800 V, 3 µF). Recombinant embryonic stem cell clones expressing the neomycin gene were selected in medium supplemented with G418 over a period of 8 days. TK was used as a counter selection. Approximately 500 recombinant embryonic stem cell clones were screened for the presence of homologous recombinants by Southern blotting. DNA was digested with Nco1 and hybridized with a probe outside the construct. Positive clones were injected into C57BL/6 blastocysts using standard procedures. The neomycin selection cassette was flanked by Frt recombination sites and excised *in vivo* by crossing the mice with the FLPeR deleter stain (Farley et al., 2000).

Cre induction in *Ah-Cre/*Ascl2$^{floxed/floxed}$ mice
The Cre enzyme was induced in mice 6-12 weeks old of age by intraperitoneal injections at day 0 of 200 µl ß-napthoflavone (10 mg ml^{-1}; Sigma Aldrich) dissolved in corn oil. All mice were maintained at the Hubrecht Institute (Utrecht, The Netherlands) according to institutional guidelines.

Isolation of intestinal cells for Southern and Northern blot analysis
To isolate small intestinal epithelial cells, intestines were isolated from animals and cut longitudinally and washed in ice-cold PBS (Mg^{2+}/Ca^{2+}). The intestine was then cut into small pieces (1–2-cm long) and washed several times in ice-cold PBS (Mg^{2+}/Ca^{2+}) to remove contaminants such as faeces and hair. The intestinal pieces were incubated in 30 mM EDTA in PBS0 (lacking Mg^{2+}/Ca^{2+}) at 37 degrees followed by shaking. The released epithelium was then collected to prepare DNA or RNA by standard procedures.

Histology, immunohistochemistry and in situ hybridization
Tissues of mice were fixed in 10% formalin, paraffin embedded, and sectioned at 3-6 µm for hematoxylin/eosin (H&E) staining, immunostaining procedure as described by (Batlle et al., 2002) or *in situ* hybridization. The primary antibodies were mouse anti-Ki67 (1:100; Novacastra), rabbit anti-c-Myc (1:500; Upstate Biotechnology), mouse anti-Ascl2 (1:5), rabbit anti-Sox9 (1:600, Chemicon), rabbit anti-Caspase-3 (1:400; Cell signaling), rabbit anti-GFP (1:6000, gift from E. Cuppen). The peroxidase conjugated secondary antibodies used were Mouse or Rabbit EnVision+ (DAKO). For in situ hybridization, mouse ESTs were obtained from the IMAGE consortium or RZPD. These clones were used for *in vitro* transcription reactions to generate probes for *in situ* hybridizations. Protocols for *in vitro* transcription and *in situ* hybridizations are described elsewhere (Gregorieff et al., 2005). ß-galactosidase (LacZ) staining for visualization of Cre mediated deletion using the Rosa26-LacZ reporter mice was done as described previously (Barker et al., 2007).

Electron microscopy analysis
Tissues were fixed in 2.5% glutaraldehyde + 2.0% paraformaldehyde in cacodylate-buffer, postfixed in 1% OsO4, stained en bloc with uranylacetate, and embedded in Epon resin. The samples were examined with a Phillips CM10 microscope (Eindhoven, The Netherlands).

Generation of a stably inducible ASCL2 over expression and ASCL2 RNAi CRC cell lines
T-Rex system (Invitrogen) was used to generate an ASCL2 inducible HCT116 CRC cell line according to the protocol as described previously (van de Wetering et al., 2002). LS174T CRC cells were used to generate stable inducible ASCL2 sh-RNA cell line using the pTER system according to the protocol described previously (van de Wetering et al., 2003). The used oligonucleotides are given in Supplemental Table 3.

Chromatin immunoprecipitation (ChIP)
ChIP protocols were performed as described previously (Hatzis et al., 2008) with small modifications. Approximately 10^7 cells were crosslinked with 1% formaldehyde for 30 min at room temperature. The reaction was quenched with glycine at a final concentration of 0.125 M. The cells were successively washed with PBS, buffer B (0.25% Triton-x 100, 10 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6) and buffer C (0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6) at 4 ^{o}C for 10 min each. The cells were then resuspended in ChIP-incubation buffer (0.3% SDS, 1% Triton-x 100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6) and sheared using a BIORUPTOR sonicator (COSMO BIO CO., LTD) with 10 pulses of 30 sec each at maximum setting. The sonicated chromatin was centrifuged for 15 min and incubated for 12 hours at 4 ^{o}C with protein G beads (Upstate) which had been preincubated for 2 h with Ascl2 hybridoma supernatant (corresponding to approximately 5 µg of antibody). The beads were successively washed 2 times with buffer 1 (0.1% SDS, 0.1% DOC, 1% Triton-x100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6), 1 time with buffer 2 (0.1% SDS, 0.1% sodium deoxycholate, 1% Triton-x100, 0.5 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6), 1 time with buffer 3 (0.25 M LiCl, 0.5% sodium deoxycholate, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6), and 2 times with buffer 4 (1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6), for 5 min each at 4 ^{o}C. The precipitated chromatin was eluted by incubation of the beads with elution buffer (1% SDS, 0.1 M $NaHCO_3$) at room temperature for 30 min, de-crosslinked by incubation at 65 ^{o}C for 5-15 hours in the presence of 200 mM NaCl, extracted with phenol/chloroform and precipitated.
For sequential ChIP, the eluted chromatin was diluted with ChIP-incubation buffer without SDS to the incubation conditions of the first ChIP. Half the amount of hybridoma supernatant was added to the second ChIP and processed as the first.

Quantitative PCR (qPCR)
ChIP samples and cDNA samples generated from GFP sorted *Lgr5-EGFP-ires-CreERT2* RNA were analyzed with qPCR performed as described previously (Hatzis et al., 2008). The used primer pairs are given in Supplemental Table 3. ChIP values were normalized as a percentage of input. ChIP enrichment was expressed as fold over respective values for control regions (i.e., exon 2 of the non expressed myoglobin gene). Based on ASCL2 occupancy values over a number of such negative control regions, we defined as positive those regions whose change in occupancy over the control region was greater than threefold. Values for the cDNA samples are shown as ratios of values in GFP^{hi} divided by GFP^{lo} values after normalization as % of β-actin or B2M housekeeping genes. For Supplementary Figure 1H values are normalized with B2M.

Reporter assays
SOX9 enhancer (oligonucleotides used for cloning are given in Suppl. Table 3) encompassing about 1 kb of genomic sequence around the ASCL2 binding region were amplified by PCR from human genomic DNA and cloned in pGL4.10 (Invitrogen) containing a minimal TATA box in front of the firefly luciferase gene. The reporter was transfected with P-PEI (Poly Sciences) in inducible LS174T ASCL2 RNAi and inducible HCT116 ASCL2 CRC cell lines. Cells were plated in 24 wells and transfected with 500 ng reporter. Renilla luciferase was used as a transfection control. Cells were induced with doxyxycline for 24 hr and reporter activity was measured using the dual luciferase reporter assay system (Promega).

Acknowledgments
We would like to thank Robert Vries, Jeroen Korving, Miranda Cozijnsen, Stieneke van den Brink and Eveline Steine for technical help. Maarten Van Lohuizen is thanked for providing Bmi1 knockout tissues.

References
Barker, N., van de Wetering, M., and Clevers, H. (2008). The intestinal stem cell. Genes Dev *22*, 1856-1864.
Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J., and Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.
Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F.*, et al.* (2007). Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol *178*, 635-648.
Batlle, E., Henderson, J. T., Beghtel, H., van den Born, M. M., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T., and Clevers, H. (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell *111*, 251-263.
Bjerknes, M., and Cheng, H. (1981a). The stem-cell zone of the small intestinal epithelium. I. Evidence from Paneth cells in the adult mouse. Am J Anat *160*, 51-63.

Bjerknes, M., and Cheng, H. (1981b). The stem-cell zone of the small intestinal epithelium. III. Evidence from columnar, enteroendocrine, and mucous cells in the adult mouse. Am J Anat *160*, 77-91.
Bjerknes, M., and Cheng, H. (1999). Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology *116*, 7-14.
Blache, P., van de Wetering, M., Duluc, I., Domon, C., Berta, P., Freund, J. N., Clevers, H., and Jay, P. (2004). SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. J Cell Biol *166*, 37-47.
Cabrera, C. V., and Alonso, M. C. (1991). Transcriptional activation by heterodimers of the achaete-scute and daughterless gene products of Drosophila. Embo J *10*, 2965-2973.
Cairnie, A. B., and Millen, B. H. (1975). Fission of crypts in the small intestine of the irradiated mouse. Cell Tissue Kinet *8*, 189-196.
Calleja, M., Renaud, O., Usui, K., Pistillo, D., Morata, G., and Simpson, P. (2002). How to pattern an epithelium: lessons from achaete-scute regulation on the notum of Drosophila. Gene *292*, 1-12.
Caudy, M., Vassin, H., Brand, M., Tuma, R., Jan, L. Y., and Jan, Y. N. (1988). daughterless, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. Cell *55*, 1061-1067.
Cheng, H., and Leblond, C. P. (1974a). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. Am J Anat *141*, 461-479.
Cheng, H., and Leblond, C. P. (1974b). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat *141*, 537-561.
Farley, F. W., Soriano, P., Steffen, L. S., and Dymecki, S. M. (2000). Widespread recombinase expression using FLPeR (flipper) mice. Genesis *28*, 106-110.
Fevr, T., Robine, S., Louvard, D., and Huelsken, J. (2007). Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. Mol Cell Biol *27*, 7551-7559.
Gregorieff, A., and Clevers, H. (2005). Wnt signaling in the intestinal epithelium: from endoderm to cancer. Genes Dev *19*, 877-890.
Gregorieff, A., Pinto, D., Begthel, H., Destree, O., Kielman, M., and Clevers, H. (2005). Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology *129*, 626-638.
Guillemot, F., Caspary, T., Tilghman, S. M., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Anderson, D. J., Joyner, A. L., Rossant, J., and Nagy, A. (1995). Genomic imprinting of Mash2, a mouse gene required for trophoblast development. Nat Genet *9*, 235-242.
Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., and Joyner, A. L. (1994). Essential role of Mash-2 in extraembryonic development. Nature *371*, 333-336.
Hatzis, P., van der Flier, L. G., van Driel, M. A., Guryev, V., Nielsen, F., Denissov, S., Nijman, I. J., Koster, J., Santo, E. E., Welboren, W.*, et al.* (2008). Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. Mol Cell Biol *28*, 2732-2744.
Hughes, M., Dobric, N., Scott, I. C., Su, L., Starovic, M., St-Pierre, B., Egan, S. E., Kingdom, J. C., and Cross, J. C. (2004). The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells. Dev Biol *271*, 26-37.
Ireland, H., Kemp, R., Houghton, C., Howard, L., Clarke, A. R., Sansom, O. J., and Winton, D. J. (2004). Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of beta-catenin. Gastroenterology *126*, 1236-1246.
Johnson, J. E., Birren, S. J., and Anderson, D. J. (1990). Two rat homologues of Drosophila achaete-scute specifically expressed in neuronal precursors. Nature *346*, 858-861.
Johnson, J. E., Birren, S. J., Saito, T., and Anderson, D. J. (1992). DNA binding and transcriptional regulatory activity of mammalian achaete-scute homologous (MASH) proteins revealed by interaction with a muscle-specific enhancer. Proc Natl Acad Sci U S A *89*, 3596-3600.
Jubb, A. M., Chalasani, S., Frantz, G. D., Smits, R., Grabsch, H. I., Kavi, V., Maughan, N. J., Hillan, K. J., Quirke, P., and Koeppen, H. (2006). Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. Oncogene *25*, 3445-3457.
Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet *19*, 379-383.
Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science *275*, 1784-1787.
Kosinski, C., Li, V. S., Chan, A. S., Zhang, J., Ho, C., Tsui, W. Y., Chan, T. L., Mifflin, R. C., Powell, D. W., Yuen, S. T.*, et al.* (2007). Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. Proc Natl Acad Sci U S A *104*, 15418-15423.
Kuhnert, F., Davis, C. R., Wang, H. T., Chu, P., Lee, M., Yuan, J., Nusse, R., and Kuo, C. J. (2004). Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A *101*, 266-271.

Mori-Akiyama, Y., van den Born, M., van Es, J. H., Hamilton, S. R., Adams, H. P., Zhang, J., Clevers, H., and de Crombrugghe, B. (2007). SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology *133*, 539-546.
Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K. W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science *275*, 1787-1790.
Muncan, V., Sansom, O. J., Tertoolen, L., Phesse, T. J., Begthel, H., Sancho, E., Cole, A. M., Gregorieff, A., de Alboran, I. M., Clevers, H., and Clarke, A. R. (2006). Rapid loss of intestinal crypts upon conditional deletion of the Wnt/Tcf-4 target gene c-Myc. Mol Cell Biol *26*, 8418-8426.
Phillips, R. G., and Whittle, J. R. (1993). wingless expression mediates determination of peripheral nervous system elements in late stages of Drosophila wing disc development. Development *118*, 427-438.
Pinto, D., Gregorieff, A., Begthel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev *17*, 1709-1713.
Pinto, D., Robine, S., Jaisser, F., El Marjou, F. E., and Louvard, D. (1999). Regulatory sequences of the mouse villin gene that efficiently drive transgenic expression in immature and differentiated epithelial cells of small and large intestines. J Biol Chem *274*, 6476-6482.
Potten, C. S. (1977). Extreme sensitivity of some intestinal crypt cells to X and gamma irradiation. Nature *269*, 518-521.
Potten, C. S., Kovacs, L., and Hamilton, E. (1974). Continuous labelling studies on mouse skin and intestine. Cell Tissue Kinet *7*, 271-283.
Reya, T., and Clevers, H. (2005). Wnt signalling in stem cells and cancer. Nature *434*, 843-850.
Riccio, O., van Gijn, M. E., Bezdek, A. C., Pellegrinet, L., van Es, J. H., Zimber-Strobl, U., Strobl, L. J., Honjo, T., Clevers, H., and Radtke, F. (2008). Loss of intestinal crypt progenitor cells owing to inactivation of both Notch1 and Notch2 is accompanied by derepression of CDK inhibitors p27(Kip1) and p57(Kip2). EMBO Rep *9*, 377-383.
Sangiorgi, E., and Capecchi, M. R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet *40*, 915-920.
Sansom, O. J., Reed, K. R., Hayes, A. J., Ireland, H., Brinkmann, H., Newton, I. P., Batlle, E., Simon-Assmann, P., Clevers, H., Nathke, I. S.*, et al.* (2004). Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. Genes Dev *18*, 1385-1390.
Sato, H., Vries, R., Snippert, H., van de Wetering, M., Barker, N., van Es, J. H., Abo, A., Kujala, P., Peters, P. J., and Clevers, H. (Submitted). Single intestinal stem cells build crypt-villus structures in vitro in the absence of a cellular niche.
Scott, I. C., Anson-Cartwright, L., Riley, P., Reda, D., and Cross, J. C. (2000). The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Mol Cell Biol *20*, 530-541.
Simpson, P. (1990). Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of Drosophila. Development *109*, 509-519.
Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.
Stepan, H., Marqwardt, W., Kuhn, Y., Hockel, M., Schultheiss, H. P., and Walther, T. (2003). Structure and regulation of the murine Mash2 gene. Biol Reprod *68*, 40-44.
van de Wetering, M., Oving, I., Muncan, V., Pon Fong, M. T., Brantjes, H., van Leenen, D., Holstege, F. C., Brummelkamp, T. R., Agami, R., and Clevers, H. (2003). Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO Rep *4*, 609-615.
van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P.*, et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.
Van der Flier, L. G., Sabates-Bellver, J., Oving, I., Haegebarth, A., De Palo, M., Anti, M., Van Gijn, M. E., Suijkerbuijk, S., Van de Wetering, M., Marra, G., and Clevers, H. (2007). The Intestinal Wnt/TCF Signature. Gastroenterology *132*, 628-632.
van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M., and et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes Dev *8*, 757-769.
van Es, J. H., Jay, P., Gregorieff, A., van Gijn, M. E., Jonkheer, S., Hatzis, P., Thiele, A., van den Born, M., Begthel, H., Brabletz, T.*, et al.* (2005a). Wnt signalling induces maturation of Paneth cells in intestinal crypts. Nat Cell Biol *7*, 381-386.
van Es, J. H., van Gijn, M. E., Riccio, O., van den Born, M., Vooijs, M., Begthel, H., Cozijnsen, M., Robine, S., Winton, D. J., Radtke, F., and Clevers, H. (2005b). Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. Nature *435*, 959-963.
Zhang, J., Liu, W. L., Tang, D. C., Chen, L., Wang, M., Pack, S. D., Zhuang, Z., and Rodgers, G. P. (2002). Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. Gene *283*, 83-93.

Supplemental data

Supplementary Table 1

GeneName	RefSeq ID	Lgr5 log2 AVG	Lgr5 q-value (%)
Lgr5	NM_010195	2.54	0.45
Tnfrsf19	NM_013869	2.23	0.45
Fstl1	NM_008047	1.70	4.07
Nkd1	NM_027280	1.65	4.07
Sp5	NM_022435	1.60	4.07
Arl4c	NM_177305	1.59	4.07
Rgmb	BC096024	1.54	4.07
1190003M12Rik	NM_026860	1.53	4.07
Acot1	NM_012006	1.51	4.07
Pdlim2	NM_145978	1.50	4.07
Slc14a1	NM_028122	1.49	4.07
NAP027049-1	NAP027049-1	1.48	4.07
Gpr30	NM_029771	1.48	4.07
2810004A10Rik	NM_027265	1.43	4.07
Lamb3	NM_008484	1.39	4.07
Ephb3	NM_010143	1.38	4.07
1190002H23Rik	NM_025427	1.36	4.07
Nrn1	NM_153529	1.35	4.07
E030025L21Rik	NM_207531	1.33	4.07
Pdgfa	NM_008808	1.32	4.07
Cyr61*	NM_010516	1.30	4.47
Plekhb1	NM_013746	1.29	4.07
Mia1	NM_019394	1.28	4.07
Nav1	NM_173437	1.27	4.07
Ces3	NM_053200	1.27	4.07
Pls3	NM_145629	1.26	4.47
Rec8L1	NM_020002	1.25	4.10
Slco3a1	NM_023908	1.21	4.42
Adora1	NM_001008533	1.21	4.47
Ptpro	NM_011216	1.19	4.47
ENSMUST00000064814	ENSMUST00000064814	1.16	4.71
BC107410	BC107410	1.16	5.82
Cd44	NM_009851	1.15	4.71
Nr2e3	NM_013708	1.14	4.71
Scn2b	NM_001014761	1.14	4.71
Rnf43	NM_172448	1.14	4.47
Sfxn3	NM_053197	1.14	4.71
Kif12	NM_010616	1.12	4.71
Sox9	NM_011448	1.12	4.71
Sectm1b	NM_026907	1.11	4.71
E030011O05Rik*	AK086536	1.11	4.71
Utrn	NM_011682	1.11	4.71
Igfbp4*	NM_010517	1.10	4.71
Sox6	NM_011445	1.09	4.71
D430015B01Rik	NM_153574	1.09	4.71
Ddit4	NM_029083	1.09	5.82
Dgkg	NM_138650	1.08	4.71
Shroom1	NM_027917	1.08	4.71
6330505N24Rik	NM_001033301	1.08	4.71
AI173486	NM_178928	1.07	4.71
1300002K09Rik	NM_028788	1.04	4.71
Parc	AK019026	1.04	4.71
Jun	NM_010591	1.03	4.71
Hunk	NM_015755	1.03	4.71
Olfm4	NM_001030294	1.02	5.82
Cdk6	AK030810	1.01	5.82
TC1722686	TC1722686	1.01	4.71
Zfp341	NM_199304	1.01	4.71
Socs2	AK033206	1.00	4.71
Sorcs2	NM_030889	1.00	4.71
Soat1	NM_009230	0.99	5.82
Acsm3	NM_212441	0.99	5.82
Fgfr4	NM_008011	0.98	5.82
Hook1	AK049897	0.97	5.82
Psrc1	NM_019976	0.97	5.82
Aqp1	NM_007472	0.97	5.82
Tns3	XM_109868	0.97	5.82
Arid5b	NM_023598	0.97	5.82

Slc23a3	NM_194333	0.96	5.82
H2-Aa	NM_010378	0.95	5.82
Cdc42ep1	NM_027219	0.95	5.82
BC057022	NM_001004180	0.94	5.82
Gas6	NM_019521	0.94	5.82
Sema7a	NM_011352	0.94	5.82
2610042L04Rik*	AK084071	0.93	6.54
Ell3	NM_145973	0.92	5.82
Tia1	NM_011585	0.92	5.82
Slc30a2	AK031425	0.91	5.82
AI428936	NM_153577	0.91	5.82
CB588406	CB588406	0.90	5.82
Phlda1	NM_009344	0.90	5.82
App	NM_007471	0.89	5.82
Emp2	NM_007929	0.88	5.82
Cited4	NM_019563	0.88	6.54
Dbp	NM_016974	0.87	5.82
Rnf32	NM_021470	0.87	6.54
Slc12a2	NM_009194	0.87	5.82
Krt23	NM_033373	0.87	6.54
Immp2l	NM_053122	0.87	6.54
1500012F01Rik	AK005231	0.86	6.54
Picalm	NM_146194	0.85	6.54
Cfi	NM_007686	0.85	6.54
H2-Eb1	X52641	0.84	7.07
Prelp	NM_054077	0.83	7.07
Fgfrl1	NM_054071	0.83	6.54
B930095G15Rik	AK084997	0.82	7.07
Trim24	NM_145076	0.82	7.07
Aqp4	NM_009700	0.82	7.07
Ptprd	NM_011211	0.82	6.54
Cachd1	XM_991907	0.82	7.07
Zbtb12	NM_198886	0.81	7.07
Lrig1	NM_008377	0.81	7.07
Notch1	Z11886	0.81	7.07
Chmp4c	AK033502	0.81	8.97
Wipi1	NM_145940	0.81	8.07
Zfp12	NM_177681	0.81	7.07
Ets2	NM_011809	0.80	7.07
2210407C18Rik	NM_144544	0.80	8.07
Zfp503	NM_145459	0.80	7.07
Tmem46	NM_145463	0.80	7.07
Prss23	NM_029614	0.79	7.07
Clca4	NM_139148	0.78	8.07
Smarcd3*	NM_025891	0.78	8.97
Sfpq	XM_994784	0.78	8.97
Adfp	NM_007408	0.78	7.07
Rpl22	NM_009079	0.78	7.07
Pik3r1	NM_001077495	0.77	8.07
Crlf1	NM_018827	0.77	8.97
Tacc1	NM_177089	0.77	8.97
Pabpc1	AK005009	0.77	8.97
Slc16a11	NM_153081	0.76	7.07
Adra2a	NM_007417	0.76	8.07
NAP019557-001	NAP019557-001	0.76	8.97
Zc3h6*	ENSMUST00000056897	0.76	8.07
Pacsin3	NM_030880	0.76	8.97
Wwp1	NM_177327	0.76	8.07
Dtx4	NM_172442	0.75	8.07
Tlr2	NM_011905	0.75	8.07
Sord	NM_146126	0.75	8.97
Ascl2	NM_008554	0.75	8.97
Cdca7	NM_025866	0.74	8.97
Pla2g5	NM_011110	0.73	8.97
Ches1	NM_183186	0.72	8.97
ENSMUST00000050697	ENSMUST00000050697	0.72	8.97

Small intestinal stem cell signature based on Lgr5 expression. GFP-positive epithelial cells from pure crypt preparations of *Lgr5-EGFP-ires-CreERT2* mice were isolated using FACS sorting. FACS analysis distinguished two populations, GFP^{hi} and GFP^{lo} cells corresponding to the CBC cells and their immediate transit-amplifying daughters, respectively. In order to identify novel stem cell genes, mRNA samples of the two populations were subjected to comparative gene expression profiling using Agilent microarray analysis. Genes which were previously identified as Wnt targets in colorectal cancer cell lines (Van der Flier et al., 2007) are indicated in gray. The 6 genes indicated with a * are not expressed in the Ascl2 arrays and for that reason not used in the comparison between stem cell genes and Ascl2 target genes as described in Figure 7A.

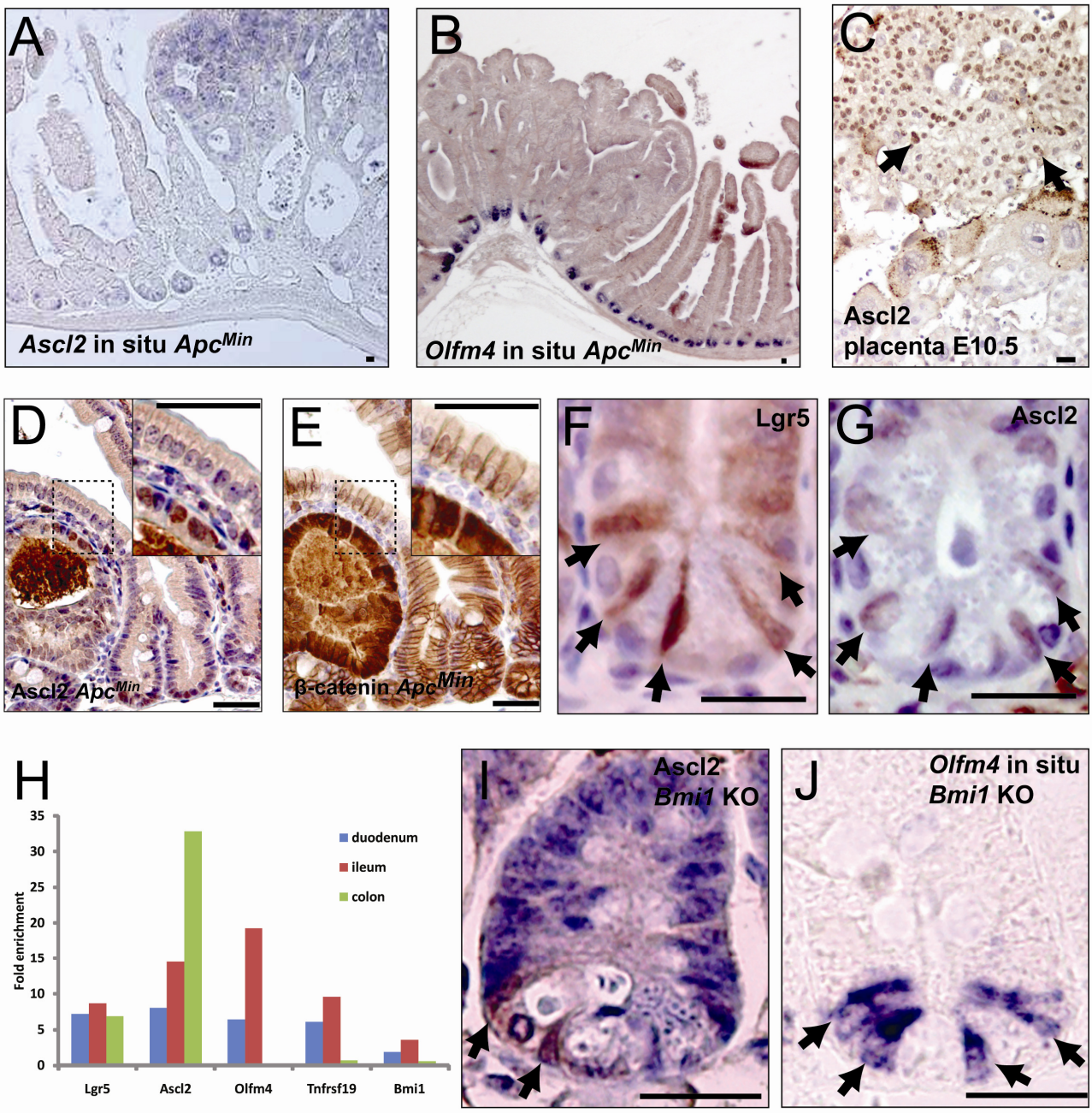
Supplementary table 2

GeneName	RefSeq ID	Ascl2 log2 AVG	Ascl2 q-value (%)	Lgr5 log2 AVG	Lgr5 q-value (%)
1190003M12Rik	NM_026860	-1.63	0.00	1.53	4.07
Adfp	NM_007408	-0.72	0.28	0.78	7.07
Adora1	NM_001008533	-0.62	1.27	1.21	4.47
Adra2a	NM_007417	-0.99	0.61	0.76	8.07
AI173486	NM_178928	-0.72	5.08	1.06	4.71
AI428936	NM_153577	-0.78	0.73	0.91	5.82
Aqp4	NM_009700	-0.92	1.06	0.82	7.07
Ascl2	NM_008554	-1.76	0.00	0.75	8.97
BC057022	NM_001004180	-0.74	0.28	0.94	5.82
Cdc42ep1	NM_027219	-0.73	2.50	0.95	5.82
Cdk6	AK030810	-0.79	1.54	1.01	5.82
Ces3	NM_053200	-0.97	0.86	1.27	4.07
Cfi	NM_007686	-2.08	2.05	0.85	6.54
Crlf1	NM_018827	-0.65	2.05	0.77	8.97
E030025L21Rik	NM_207531	-1.65	0.73	1.33	4.07
Ell3	NM_145973	-0.64	1.06	0.92	5.82
ENSMUST00000064814	ENSMUST00000064814	-1.27	0.50	1.16	4.71
Ephb3	NM_010143	-0.84	1.54	1.38	4.07
Kif12	NM_010616	-1.73	0.00	1.12	4.71
Lgr5	NM_010195	-0.85	0.47	2.54	0.45
Mia1	NM_019394	-1.10	0.00	1.28	4.07
NAP027049-1	NAP027049-1	-1.70	0.00	1.48	4.07
Nr2e3	NM_013708	-1.72	0.28	1.14	4.71
Olfm4	NM_001030294	-2.31	0.00	1.02	5.82
Psrc1	NM_019976	-0.96	3.19	0.97	5.82
Ptpro	NM_011216	-0.58	0.28	1.19	4.47
Slc12a2	NM_009194	-0.66	0.00	0.87	5.82
Slc14a1	NM_028122	-0.72	2.50	1.41	4.07
Slco3a1	NM_023908	-0.88	0.00	1.21	4.42
Soat1	NM_009230	-0.63	1.54	0.99	5.82
Sox9	NM_011448	-0.59	5.08	1.07	5.82
Tnfrsf19	NM_013869	-0.89	1.54	2.23	0.45
Zbtb12	NM_198886	-0.61	0.86	0.81	7.07

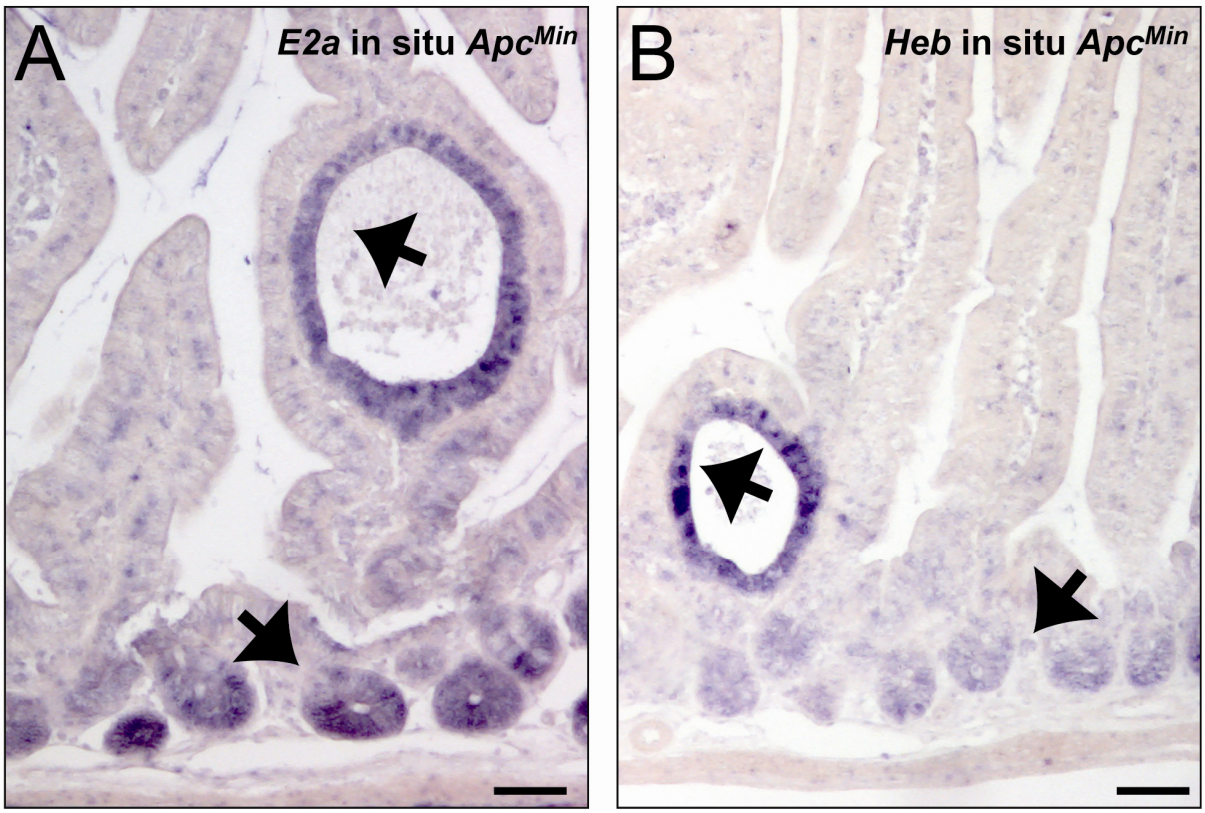
Supplemental Table 2 Intestinal Stem Cell Transcriptome regulated by Ascl2
In order to identify stem cell genes that are regulated by Ascl2 (Figure 7A), we compared the genes of the intestinal stem cell signature (Supplemental Table 1) with down regulated genes upon Ascl2 knock down. Ascl2 target genes have been identified through comparative gene profiling of RNA samples from isolated intestinal epithelium of *Ah-Cre/Ascl2*$^{floxed/floxed}$ animals and *Ah-Cre/Ascl2*$^{floxed/wt}$ control animals at day 3 and 5 days PI.

Supplementary Figure 1 Ascl2 and *Olfm4* expression
(A) *In situ* hybridization for *Ascl2* on *Apc*^{min} tissue demonstrates crypt-specific staining in wild type crypts and in adenoma tissue.
(B) *In situ* hybridization for *Olfm4* on *Apc*^{min} tissue showing crypt-specific staining in wild type crypts and no staining in adenoma tissue.
(C) Monoclonal Ascl2 antibody staining of E10.5 placenta tissue results in a clear nuclear signal (arrows) in the spongiotrophoblast layer.
(D) Monoclonal Ascl2 antibody staining on *Apc*^{min} tissue results in positive nuclei in wild type crypts and in early adenoma tissue.
(E) The staining observed in D colocalizes with nuclear β-catenin underscoring that Ascl2 is a Wnt regulated gene in the intestinal epithelium and in colorectal cancer.
(F, G) Serial sections of *Lgr5-EGFP-ires-CreERT2* knock-in mice stained with a GFP antibody (F) and the Ascl2 antibody (G) results in staining of the same CBC cells (arrows).
(H) RT-PCR confirmation of enrichment of Lgr5, Ascl2, Olfm4, Tnfrsf19 and Bmi1 in GFP sorted duodenum, ileum and colon epithelial cells from *Lgr5-EGFP-ires-CreERT2* mice. Values represent enrichment of respective gene product in GFP^{hi} vs. GFP^{lo} cells; gene expression is normalized with the housekeeping gene B2M in these fractions. Normalization with β-actin yields essentially identical results. Experiments have been performed three times. A representative experiment is shown.
(I) Monoclonal Ascl2 antibody staining on 4 week- *Bmi1* knock-out tissue still results in CBC-specific staining in crypts (arrows).
(J) In situ hybridization for *Olfm4* on *Bmi1* knock-out tissue still shows CBC-specific staining in crypts (arrows).
Size bars represent 20 μm.

Supplementary Figure 1

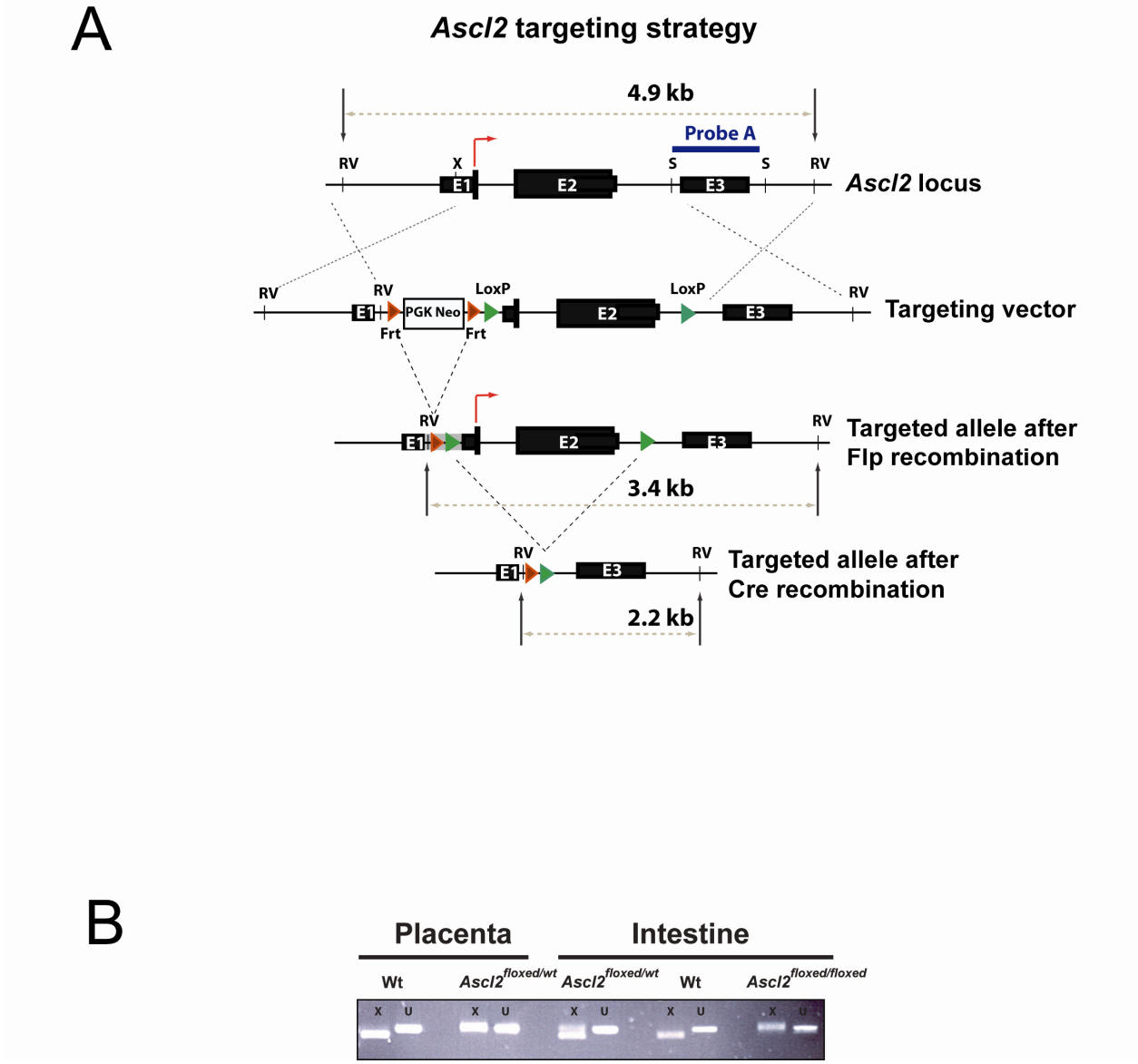


Supplementary Figure 2



Supplementary Figure 2 Expression of Ascl2 binding partners in the intestine
(A) In situ hybridization for *E2a* on *Apc*^{min} tissue showing crypt-specific staining in wild type crypts and in adenoma tissue (arrows).
(B) In situ hybridization for *Heb* on *Apc*^{min} tissue showing crypt-specific staining in wild type crypts and in adenoma tissue (arrows). Size bars represent 50 µm.

Supplementary Figure 3

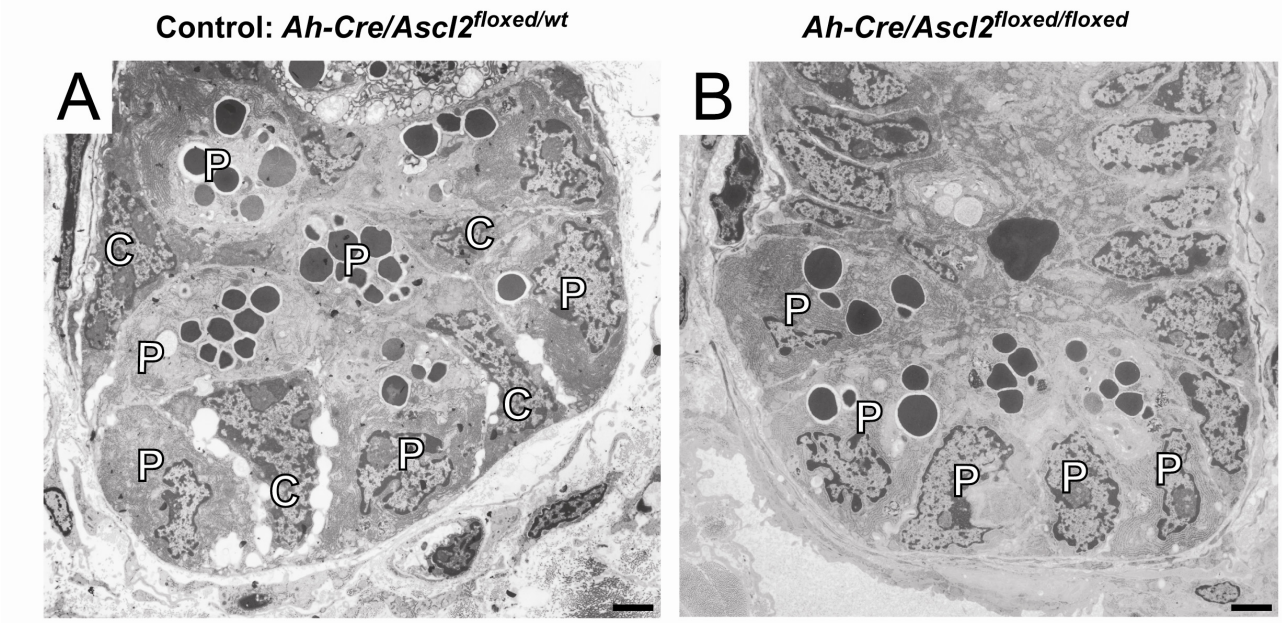


Supplementary Figure 3 Targeting strategy for conditional Ascl2 deletion; loss of paternal Ascl2 imprinting in adult intestine

(A) Targeting strategy for conditional *Ascl2* allele. The three *Ascl2* exons are indicated as black boxes E1-E3. We introduced a LoxP site into the 5' UTR of the gene. A second LoxP site was introduced between exon II and the untranslated third exon. The neomycin cassete was surrounded by Frt sites. Indicated restriction sites: X = Xho1, S = Sac1, RV = EcoRV. For EcoRV restriction analysis the fragments with their expected sizes are indicated as dotted gray lines. The internal probe A, shown as a blue bar, is used for (EcoRV) Southern blotting analysis to distinguish between wild type, floxed and floxed recombined alleles.

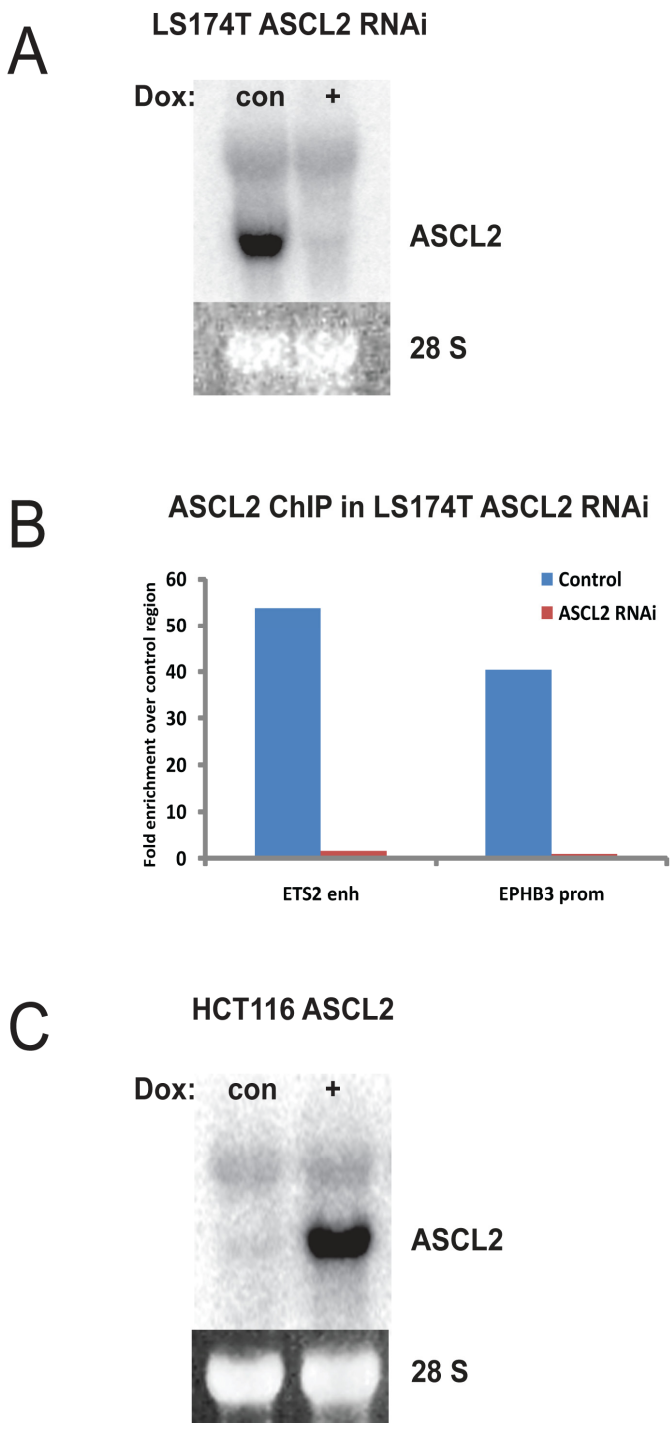
(B) Restriction digest showing that *Ascl2* is not imprinted in the intestine. Through the introduction of the FRT and LoxP sites in the 5' UTR of the gene, an Xho site is destroyed in the conditional animals. To determine if *Ascl2* is imprinted in the intestine we analyzed material from wild type, $Ascl2^{floxed/wt}$ and $Ascl2^{floxed/floxed}$ intestines and wild type and $Ascl2^{floxed/wt}$ placenta material. Xho restriction digest (X) and not treated material (U) of the PCR product surrounding the Xho site in the 5' UTR shows that the wild type placenta cDNA is completely digested while the heterozygous floxed placenta (allele inherited from the mother) cDNA cannot be digested by Xho, confirming the imprinting of Ascl2 in the mouse placenta. The partial digest of the cDNA from heterozygous floxed intestine (allele inherited from the mother) proves that Ascl2 is bi-allelically expressed in the intestine. This restriction analysis was also confirmed through subcloning followed by sequencing (data not shown).

Supplementary Figure 4



Supplementary Figure 4 Electron microscopy of crypt bottoms 5 day after Ascl2 deletion
(A) CBC cells and Paneth cells are intermingled in wild type crypts from *Ah-Cre/Ascl2*$^{floxed/wt}$ animals. Typically, Paneth cells (extensive cytoplasm and granules; P) are flanked by CBC cells (scarce cytoplasm; C) at both sides.
(B) In mutant crypts from *Ah-Cre/Ascl2*$^{floxed/floxed}$ animal 5 days PI, CBC cells are conspicuously absent. Note five Paneth cells without intermingled CBC cells.

Supplementary Figure 5



Supplementary Figure 5 ASCL2 expression in CRC cell lines and ASCL2 ChIP
(A) Northern blot showing ASCL2 knock-down in LS174T CRC cells with inducible stably integrated RNAi directed against ASCL2, after 24 hr doxycycline induction.
(B) ASCL2 ChIP in LS174T ASCL2 RNAi cell line to prove specific ASCL2 binding to indicated regions in control cells (blue bars) and cells after induction of RNAi against ASCL2 (red bars). ASCL2 binding to indicated regions is expressed as relative enrichment of the respective qPCR product over the pPCR product of the non-bound exon2 of the myoglobin gene.
(C) Northern blot showing up regulation of ASCL2 expression in HCT116 CRC cell line with inducible ASCL2, after 24 hr doxycycline induction.

Chapter 6

Mouse models to study intestinal stem cells

Mouse models to study intestinal stem cells

Laurens G. van der Flier, Hugo J. Snippert, Jeroen Korving, Johan H. van Es and Hans Clevers

Hubrecht Institute and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

Abstract

The intestinal epithelium is a very attractive tissue to study mammalian stem cell biology. The epithelium is rapidly self-renewing and shows clear compartmentalization of proliferating and differentiated cell types. The recent identification of a definitive intestinal stem cell marker, Lgr5/Gpr49, makes it possible to visualize and distinguish intestinal stem cells from their transit amplifying progeny. The use of genetic mouse models allows us to manipulate and/or visualize specific gene expression and to study the function of these genes. Here we describe the generation of two new mouse knock-in models, which will be used for the further characterization of intestinal stem cells.

Introduction

The intestine is composed of crypts and villi covered with a specialized simple epithelium. This epithelium performs the primary functions of digestion, absorption and forms a barrier against luminal pathogens. Crypts contain proliferating stem cells and their transit amplifying daughter cells. When cells exit the crypts and enter the villi, they terminally differentiate into one of three specialized epithelial cell types: enterocytes, goblet cells and enteroendocrine cells. As an unusual cell type, Paneth cells escape the crypt-villus flow by migrating to crypt bottoms where they live for several weeks. With the exception of stem cells and Paneth cells, the intestinal epithelium is renewed approximately every 5 days [1].
The primary driving force behind the proliferation of epithelial cells in the intestinal crypts is the Wnt signaling pathway. Mice that are mutant for the intestine-specific Tcf4 transcription factor, fail to establish proliferative crypts during late gestation [2], while conditional deletion of ß-catenin [3, 4] as well as transgenic expressing of the secreted Dickkopf-1 Wnt inhibitor [5, 6] leads to disappearance of proliferative crypts in adult mice. Moreover, malignant transformation of intestinal epithelium is almost invariably initiated by activating Wnt pathway mutations [7, 8]. Because of the intimate connection between Wnt signaling and intestinal biology, we have attempted to unravel the Tcf4 target gene program activated by this pathway in crypts and colorectal tumors [9, 10]. Through this approach, we have identified the *Lgr5* gene as a marker for CBC cells [11]. Each crypt harbours around six of these small, cycling cells that are intermingled with Paneth cells at crypt bottoms. Using Cre-mediated genetic tracing, we have demonstrated that CBC cells represent long-lived, multipotent stem cells [11], as predicted originally by Leblond and colleagues [12-16].
In search for potential regulators of the intestinal stem cell fate, we have determined gene expression profiles of CBC cells of the small intestine, sorted on the basis of Lgr5-GFP expression [17]. *Olfm4* was identified as one of the stem cell specific genes found in the small intestine. To study the function of this gene in more detail, we generated a knock-in mouse. We introduced a cDNA encoding for a fluorescent Cherry fused to a herpes simplex virus type 1 thymidine kinase (tk), in frame with the ATG of the first exon of the *Olfm4* gene. Besides the analysis of the homozygous null-allele, we expect that we can use this allele both for visualization of intestinal stem cells and to inducibly deplete these cells through ganciclovir injections.
A second mouse model that we have generated is a Cre-activatable multi-color reporter strain using the ubiquitously active *Rosa26* locus. Cre-mediated excision of the roadblock sequence in this reporter strain will genetically mark individual cells with one of multiple, distinct colors. The stochastic choice between four fluorescent proteins after a Cre pulse enables to label individual cells uniquely. We will cross this multi-color reporter strain with the stem cell specific *Lgr5-CreERT2* knock-in allele. Activation of this stem cell specific Cre through tamoxifen injections will facilitate lineage tracing experiments whereby the various stem cells within a crypt will be labeled with different fluorescent clonal marks. This gives us the opportunity to study the dynamics of the pool of stem cells located within single crypts of the intestine.

Results and Discussion

Olfm4 is an intestinal stem cell marker

Recently, we described the first definitive adult intestinal stem cell marker gene; Lgr5/Gpr49. Lgr5 is an orphan G protein–coupled receptor that we originally identified in Wnt target gene expression arrays [9, 10]. Lgr5 expression in the intestine is restricted to CBC cells. Lineage tracing experiments using an *Lgr5-EGFP-ires-CreERT2* knock-in allele crossed with the *Rosa26RLacZ* reporter proved that Lgr5 is a marker for long-lived, pluripotent intestinal stem cells in both the small intestine and the colon [11].
In order to define a gene expression profile for $Lgr5^+$ intestinal stem cells, we sorted GFP-positive epithelial cells from cell suspensions prepared from freshly isolated crypts of *Lgr5-EGFP-ires-CreERT2* mice [17]. Gene expression profiling resulted in a comprehensive list of transcripts enriched in CBC cells. One of the transcripts identified was the *Olfactomedin-4* (*Olfm4*) gene. *In*

situ hybridization experiments of mouse small intestine confirmed a highly-specific and robust staining for this novel CBC cell marker [17, 18]. *OLFM4* was recently also found to be enriched in a microarray analysis of human colon crypt material [18]. OLFM4 is a secreted molecule originally cloned from human myeloblasts [19]. No clear function has been assigned to this protein. Figure 1A and B shows *in situ* hybridizations performed on human small intestine and colon, illustrating the restricted expression pattern of *Olfm4* at crypt bottoms.

***Olfm4-CherryΔtk* knock-in allele**

In order to study the function of the *Olfm4* gene, we decided to take a genetic knock-in approach. We generated a fusion protein between fluorescent Cherry and a truncated version of HSV1 thymidine kinase (Δtk). This truncated version retains tk function but does not cause male sterility [20, 21]. The CherryΔtk fusion was tested *in vitro* by transient transfection in HEK293T cells. Cherry fluorescence was observed and transfected cells died after addition of ganciclovir to the medium (data not shown). Figure 1C shows the targeting strategy for the *Olfm4* knock-in allele. We introduced the cDNA encoding for the fusion protein in frame with the ATG in the first exon of the *Olfm4* locus through homologous recombination in embryonic stem (ES) cells. We expect that the targeted allele is a functional null for Olfm4. The neomycin cassette, included for ES cell selection, was flanked by loxP sites. Figure 1D shows a Southern blot of targeted ES cell clones showing one clone with the correct homologous recombination. Clones that showed the correct homologous recombination were used for the generation of mice.

We are planning to use these mice for: (I) Study the function of the *Olfm4* gene by analyzing intestines of homozygous *Olfm4* knock-in animals. (II) Visualization and isolation of small intestinal stem cells based on fluorescent cherry expression. (III) Study the dynamics of intestinal epithelium after inducible deletion of CBC cells through glanciclovir injections.

Multi-color reporter mouse

Reporter mice, like the Cre activatable *Rosa26RLacZ*, are very powerful genetic tools for the identification of stem cells. The genetic inheritance of a clonal mark (e.g. LacZ) enables to experimentally link (Cre expressing) stem cells specific to their progeny. Recently, Livet and colleagues described the generation of Brainbow transgenes [22]. In these transgenes, Cre/loxP recombination is used to create a stochastic choice of expression between various fluorescent proteins. The authors made transgenic lines where the Brainbow constructs were placed under the control of regulatory elements from the Thy1 gene, which are known to drive expression at high levels in a variety of neuronal cell types.

In order to have a more generally usable multi-color reporter mouse, we knocked the Brainbow construct behind a roadblock sequence in the *Rosa26* locus. Figure 2A shows a schematic representation of the Cre-activatable multi-color reporter mouse that we generated. The Brainbow-2.1 construct (containing the 4 different fluorescent proteins delimited by loxP sites in opposite orientations and positioned in tandems) was cloned behind a neomycin^R cassette preceded by a loxP site. This construct was introduced under the control of the *Rosa26* promoter. Temporary Cre expression will excise the roadblock sequence and activate one of the fluorescent proteins. This is a stochastic process, because the fluorescent proteins are orientated in opposite directions in two tandems with loxP sites orientated in two orientations. Cre expression will induce random deletion of one of the two tandems while the other tandem can inverse. Thus, upon a Cre pulse one of the four fluorescent proteins will be placed under the control of the *Rosa26* promoter. Of note: a second Cre pulse can invert the remaining tandem and result in the stochastic choice between the two remaining fluorescent proteins. This feature can be used for successive clonal marking strategies.

We targeted ES cells with the multi-color construct. Correctly recombined ES cell clones were tested using a southern screen. ES cell clones that showed correct homologous recombination were tested *in vitro* by transient Cre transfection. Non-transfected control cells did not show any fluorescence. As expected, we observed patches of fluorescent cells, expressing one of the four fluorescent colors, in the transfected clones. All four different fluorescent proteins were observed, demonstrating that all four possible Cre mediated outcomes occur. Figure 2B shows the possible outcomes after Cre mediated recombination (inversion/excision) and four groups of ES cells

expressing either: nGFP, YFP, RFP, or mCFP. *In vitro* tested clones were used for the generation of mice.
We will cross this reporter strain with the stem cell specific *Lgr5-GFP-ires-CreERT2* knock-in allele [11]. This will facilitate lineage tracing experiments whereby the various stem cells within a crypt will be labeled by different clonal marks. We will cross this multi-color mouse also to homozygosity. This will increase the power of the analysis, since 10 different fluorescent outcomes (clonal marks) can be expected in these animals (Figure 2C). The analysis at different time points after temporal Cre induction will give us the opportunity to study the dynamics of the pool of stem cells located within single crypts of the intestine. Besides the analysis of the normal stem cells we will also use these animals for the analysis of cancer stem cells. According to the cancer stem cell theory, tumors are generated and maintained by a small (defined) subset of undifferentiated cells able to self-renew and differentiate into the bulk tumor population [23-25]. Recently we have shown that $Lgr5^+$ crypt stem cells are the cells of origin for intestinal adenomas [26]. We are planning to repeat these experiments in combination with the multi-color reporter. Through low tamoxifen inductions of the Lgr5 Cre we will initiate clonal adenoma formation. After a couple of weeks, a second tamoxifen induction will be used to re-label Lgr5 expressing adenoma cells. Lineage tracing of these re-marked adenoma cells will help us to show if the cancer stem cell theory is valid.

Conclusion

In this chapter, we describe the generation of two new mouse knock-in models which will allow the study of stem cell biology of the intestine. The first model will be used to study the function of the stem cell specific *Olfm4* gene. Moreover, with the generation of this knock-in allele we expect to be able to both visualize Olfm4 positive stem cells and to inducibly deplete these cells. This will allow us to see how the intestine reacts to deletion of intestinal stem cells. The second model is a modified multi-color reporter mouse that can be used to conditionally mark cells with different fluorescent clonal marks. We will breed this strain with *Lgr5-GFP-ires-CreERT2* and floxed *APC* mice. Lineage tracing experiments after tamoxifen-induced *Cre* activation in these mice will give us the opportunity to study the dynamics of (cancer) stem cells in the intestine.

Figure 1

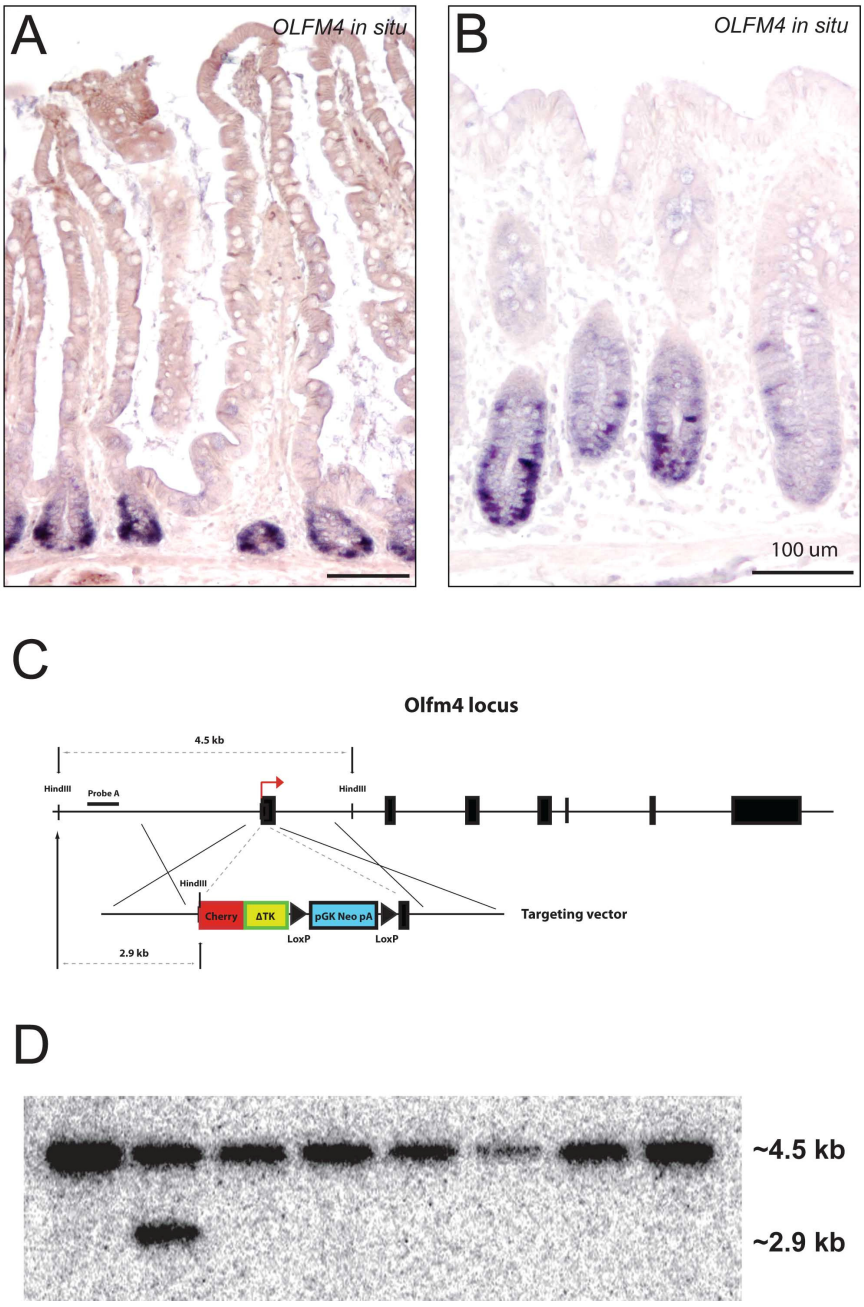


Figure 1 ***Olfm4-CherryΔtk*** **knock-in allele**
a, b, *In situ* hybridization performed on human small intestine (a) and colon (b) tissues, illustrating the expression of *OLFM4* at the crypt base.
c, Shows the targeting strategy for the generation of *Olfm4-CherryΔtk* knock-in allele. The CherryΔtk fusion is introduced in frame of the ATG in the first exon of the *Olfm4* gene. Neo, neomycin resistance cassette.
d, Southern blot analysis to screen for ES cell clones with the correct homologous recombination. ES cell DNA was digested with HindIII and hybridized with the probe A as indicacted in figure 1c.

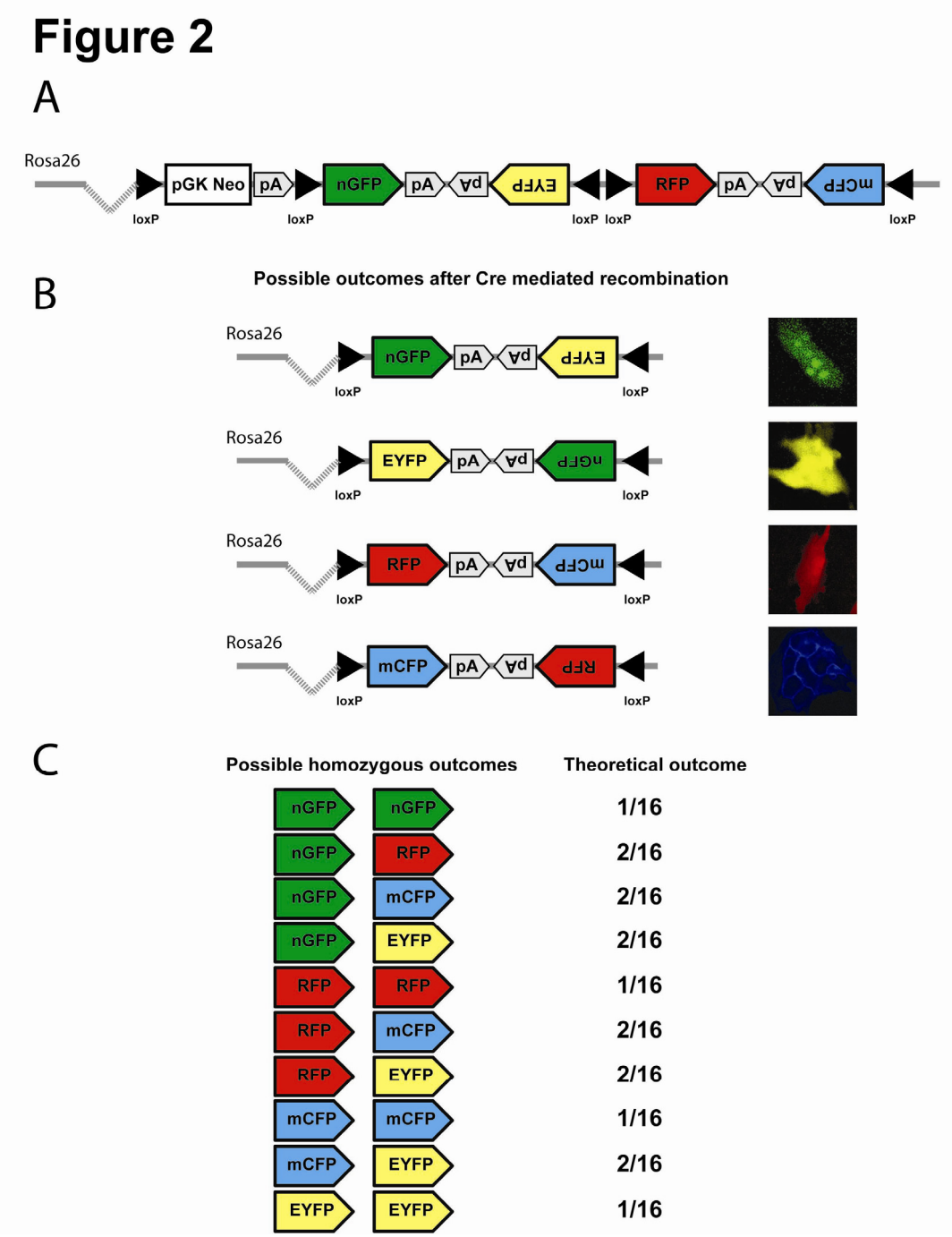


Figure 2 Multi-color mouse
a, The Cre-activatable multi-color mouse. The Brainbow-2.1 construct (containing nGFP, YFP, RFP and mCFP delimited by loxP sites in opposite orientations and positioned in tandems) was cloned behind the *Rosa26* promoter. In between a neomycin selection cassette flanked by loxP sites functions as roadblock sequence.
b, Cre mediated recombination (inversion/excision) results in a stochastic choice which of the four fluorescent proteins will be placed under the control of the *Rosa26* promoter. Targeted ES cell clones that showed the correct homologous recombination where tested *in vitro* through transient Cre transfection. Within the transfected ES cells groups of cells were observed that either expressed nGFP, YFP, RFP, or mCFP, underscoring that the four possible Cre mediated outcomes occur.
c, Homozygous multi-color mice after Cre mediated recombination have 10 different clonal outcomes. The theoretical outcome of the different color combinations is given, set out that the change of activation of the four colors within the locus is identical.

Material and Methods

OLFM4 In situ hybridization experiment

An human *OLFM4* ESTs was used to generate an *in situ* probe. Protocols for *in vitro* transcription and *in situ* hybridizations are described elsewhere [27].

Generation of *Olfm4* and multi-color knock-in mice

The Olfm4 targeting construct was generated by cloning the various components into pBluescript2 SK+ vector. A cDNA encoding for a fusion of Cherry and Δtk [28] was generated and put in frame with the ATG of the Olfm4 gene. Both flanking arms of Olfm4 were generated by high-fidelity PCR reactions from male 129/Ola-derived IB10 embryonic stem cell DNA. The multi-color construct was generated by cloning the fluorescent Brainbow 2.1 cassette [22] as NheI/HindIII fragment behind a loxP site followed by a neomycin cassette containing a stop sequence. Subsequently this fragment was cloned as Asc1/Pac1 into a vector containing Rosa26 flanking arms. The targeting constructs (100 μg) were linearized and transfected into male 129/Ola-derived IB10 embryonic stem cells by electroporation (800 V, 3 μF). Recombinant embryonic stem cell clones expressing the neomycin genes were selected in medium supplemented with G418 over a period of 8 days. Recombinant embryonic stem cell clones were screened for the presence of homologous recombinants by Southern blotting. DNA was digested with HindIII (Olfm4 construct) and EcoR1 (multi-color construct) and hybridized with probes outside the constructs. Positive clones were injected into C57BL/6 blastocysts using standard procedures. The neomycin selection cassette in the Olfm4 knock-in mouse is flanked by loxP sites and will be excised *in vivo* by crossing the mice with EIIa-CreA stain [29].

References

1. Barker, N., van de Wetering, M. & Clevers, H. The intestinal stem cell. Genes Dev 22, 1856-64 (2008).
2. Korinek, V. et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 19, 379-83 (1998).
3. 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 27, 7551-9 (2007).
4. Ireland, H. et al. Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of beta-catenin. Gastroenterology 126, 1236-46 (2004).
5. Kuhnert, F. et al. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A 101, 266-71 (2004).
6. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev 17, 1709-13 (2003).
7. Korinek, V. et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 275, 1784-7 (1997).
8. Morin, P. J. et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 275, 1787-90 (1997).
9. van de Wetering, M. et al. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 111, 241-50 (2002).
10. Van der Flier, L. G. et al. The Intestinal Wnt/TCF Signature. Gastroenterology 132, 628-32 (2007).
11. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449, 1003-7 (2007).
12. Bjerknes, M. & Cheng, H. The stem-cell zone of the small intestinal epithelium. I. Evidence from Paneth cells in the adult mouse. Am J Anat 160, 51-63 (1981).

13. Bjerknes, M. & Cheng, H. The stem-cell zone of the small intestinal epithelium. III. Evidence from columnar, enteroendocrine, and mucous cells in the adult mouse. Am J Anat 160, 77-91 (1981).
14. Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology 116, 7-14 (1999).
15. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat 141, 537-61 (1974).
16. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. Am J Anat 141, 461-79 (1974).
17. van der Flier, L. G. et al. Transcription factor Achaete Scute-like 2 (Ascl2) controls intestinal stem cell fate. (Submitted).
18. Kosinski, C. et al. Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. Proc Natl Acad Sci U S A 104, 15418-23 (2007).
19. Zhang, J. et al. Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. Gene 283, 83-93 (2002).
20. Cohen, J. L. et al. Fertile homozygous transgenic mice expressing a functional truncated herpes simplex thymidine kinase delta TK gene. Transgenic Res 7, 321-30 (1998).
21. Salomon, B. et al. A truncated herpes simplex virus thymidine kinase phosphorylates thymidine and nucleoside analogs and does not cause sterility in transgenic mice. Mol Cell Biol 15, 5322-8 (1995).
22. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 450, 56-62 (2007).
23. Dick, J. E. Breast cancer stem cells revealed. Proc Natl Acad Sci U S A 100, 3547-9 (2003).
24. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. Nature 414, 105-11 (2001).
25. Wang, J. C. & Dick, J. E. Cancer stem cells: lessons from leukemia. Trends Cell Biol 15, 494-501 (2005).
26. Barker, N. et al. Crypt stem cells as the cells-of-origin of intestinal adenomas. (Submitted).
27. Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 129, 626-38 (2005).
28. Chen, Y. T. & Bradley, A. A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells. Genesis 28, 31-5 (2000).
29. Lakso, M. et al. Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. Proc Natl Acad Sci U S A 93, 5860-5 (1996).

Chapter 7

Function of the *Regenerating* (*Reg*) gene family in the intestine

Function of the *Regenerating* (*Reg*) gene family in the intestine

Laurens G. van der Flier, Johan H. van Es, Maaike van den Born, Miranda Cozijnsen, Jeroen Korving, Victor Guryev and Hans Clevers

Hubrecht Institute and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

Abstract

The *Reg* gene family encompasses a group of highly homologous, secreted proteins containing a lectin binding domain. With the exception of Reg4, all family members are encoded in a single cluster in the genome. Physiologically, the various Reg family members are predominantly expressed in pancreas and in the intestinal epithelium. Under pathological conditions, *Reg* genes are however frequently up regulated in diverse tissues. The exact function of the Reg proteins has remained unclear. In order to study the function of the *Reg* gene family, we took a genetic loss-of-function approach in mice. The *Reg* cluster was floxed through two successive targeting rounds in embryonic stem cells, replacing the two outmost *Reg* genes by loxP sites. Independently, a conditional *Reg4* allele was generated. Through cross-breeding of these floxed animals with the *Ah-Cre* mice, we will be able to delete all *Reg* genes in an inducible manner and study their exact function.

Introduction

The *Regenerating* (*Reg*) gene family encompasses a group of highly homologous, secreted proteins containing a C-type lectin binding domain. The family consists of 5 family members in man and 7 in mouse, which can be grouped into subclasses: type I, II, III, and IV [1]. The *Reg* genes are encoded in a cluster in the genome, except *Reg4* which is located on a different chromosome [1]. *Reg1* was originally identified in a ***reg***enerating islet-derived cDNA library screen in rats [2], hence the name of this gene family. Physiologically, the various Reg family members are predominantly expressed by the pancreas and the intestinal epithelium. Under pathological conditions, Regs are frequently expressed ectopically in various other tissues. We found several *Reg* genes up regulated in a microarray expression study of prospectively collected colorectal adenomas compared with normal mucosa from the same individuals [3].

The exact function of the Reg proteins is unclear. Single knock-outs have been generated for *Reg1* and *Reg3b* [4, 5]. The observed phenotypes for these knock-outs are mild, probably because of redundancy. In order to study the function of the entire *Reg* gene family in the intestine, we took a genetic loss-of-function approach in mice. Both the genes from the *Reg* cluster and the *Reg4* gene were flanked by loxP sites facilitating possible Cre mediated recombination. To study the role of the *Reg* gene family in the intestine, we will cross the floxed *Reg* cluster and floxed *Reg4* animals with the *Ah-Cre* mouse [6]. In these mice, expression of the *Cre* recombinase is driven by the promoter of the *Cyp1a* gene, which is inducible in a number of tissues upon injection of lipophilic xenobiotic agents such as ß-Naphtoflavone.

Results

Genomic organization of the *Reg* cluster and homology between Reg proteins

The Reg proteins belong to a family of closely related, secreted proteins that contain a lectin domain. The human genome encodes 5 different Reg proteins, while the mouse genome harbors 7 family members. Both in human and mouse, the *Reg* genes are encoded in a single cluster in the genome [7, 8]. Reg4 is the only exception, being located on a different chromosome. Figure 1A shows a schematic representation of the chromosome organization of the *Reg* cluster in the mouse on chromosome 6.

The different Reg proteins are highly homologues. Figure 1B shows a phylogenetic three for the various human and mouse family members based on the comparison of their full-length protein sequences. As can be observed, the hREG3G, hREG3A, mReg3g and mReg3b cluster together. The mReg3a and mReg3d proteins also belong to this cluster, although they are less conserved. Another subfamily is formed by hREG1A, hREG1B, mReg1 and mReg2. The hREG4 and mReg4 proteins also cluster together, yet clearly represent more distant family members.

Figure 1

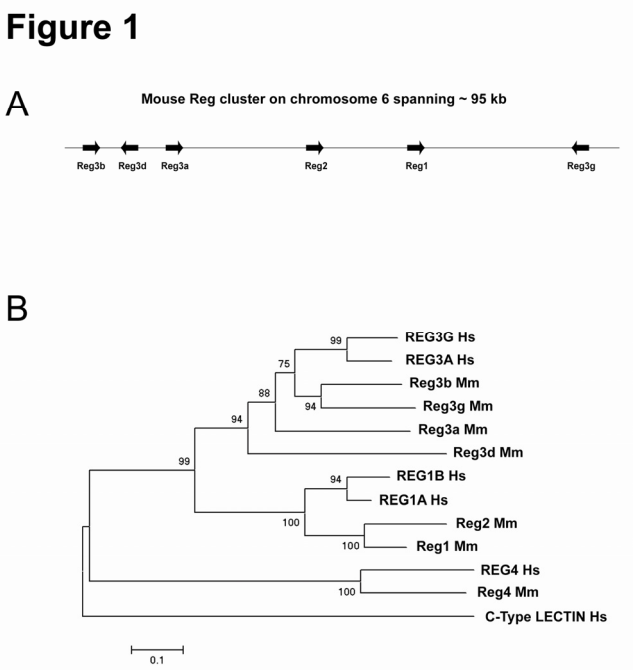


Figure 1 Genomic organization and the homology between the Reg family members
a, Schematic representation of the mouse *Reg* cluster on chromosome 6.
b, Phylogenetic tree showing the protein conservation between the five human and seven mouse Reg family members.

Expression patterns of the different *Reg* family members in the small intestine and pancreas

To study the expression pattern of the various *Reg* family members, we performed *in situ* hybridization experiments using wild-type pancreas and Apc^{min} intestines. *Reg1* is expressed in normal crypts and adenomas of the intestine (Figure 2A). In the pancreas, Reg1 expression occurs in the acinar cells, while the islets do not express *Reg1* (Figure 2B). Similarly, expression of *Reg2* is predominantly observed in the acinar cells of the pancreas, however the expression of *Reg2* in the intestine is very weak (Figure 2C and D). *Reg3a* is expressed in wild-type crypts and adenomas of the intestine and in the acinar cells of the pancreas (Figure 2E and F). The expression seen in the pancreas is, however, clearly less strong than the expression levels observed for *Reg1* and *Reg2*. *Reg3b* is expressed in the upper part of the crypts and at the crypt-villus junctions of the intestine, while adenomas are negative (Figure 2G). In the pancreas, a patchy expression pattern is observed in the acinar cells (Figure 2H). *Reg3d* expression is found at low levels in crypts, adenomas and the pancreas (Figure 2I and J). *Reg3g* shows the same expression pattern as *Reg3b* in the intestine. High level expression occurs in the upper part of the crypts and at the crypt-villus junctions, while adenomas are negative (Figure 2K). *Reg3g* is however not expressed in the pancreas (Figure 2L). *Reg4* is expressed at single cells in crypts and sporadically along the villus, while adenomas have a weak staining (Figure 2M). The pancreas is negative for *Reg4* (Figure 2N). The *in situ* hybridization experiments showed that the various *Reg* family members are not expressed in the colon (data not shown) with the exception of *Reg4* (see below).

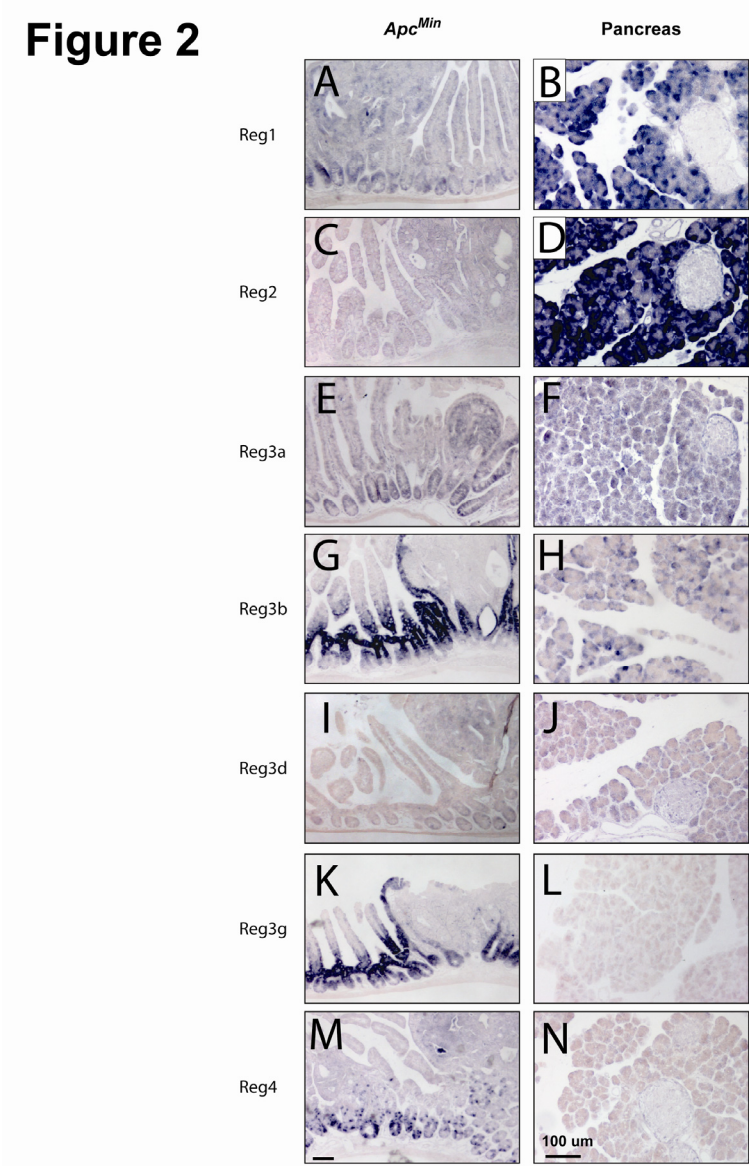


Figure 2 Expression patterns of the various *Reg* genes in the small intestine and pancreas

In situ hybridization experiments showing the expression patterns of the various *Reg* gene family members in mouse Apc^{min} tissue (**a, c, e, g, i, k and m**) and wild type pancreas (**b, d, f, h, j, l and n**).

a, b, *Reg1* is expressed in wild type crypts and adenomas of the intestine and in the acinar cells of the pancreas.

c, d, Expression of *Reg2* is predominantly observed in the acinar cells of the pancreas and very weak in the intestine.

e, f, *Reg3a* is expressed in wild type crypts and adenomas of the intestine and in the acinar cells of the pancreas.

g, h, *Reg3b* is expressed in the upper part of the crypts and at the crypt-villus junctions of the intestine, while adenomas are negative. The acinar cells of the pancreas show a patchy expression pattern for *Reg3b*.

i, j, *Reg3d* is expressed at low levels in crypts, adenomas and in the pancreas.

k, l, *Reg3g* is expressed at high levels in the upper part of the crypts and at the crypt villus junctions, while adenomas and the pancreas are negative.

m, n, *Reg4* is expressed at single cells in crypts and sporadically along the villus, adenomas have a week staining. Size bars represent 100 μm.

***Reg3b* knock-in allele**

For two of the Reg family members, single knock-out mice have been published [4, 5]. In order to study the function of the *Reg* gene family, we decided to take a knock-out approach whereby we can inducibly delete the complete Reg cluster through the introduction of loxP sites in the two outermost *Reg* genes of the cluster. Since the cluster spans ~95 kb, we targeted embryonic stem (ES) cells with two successive targeting constructs to introduce both loxP sites.

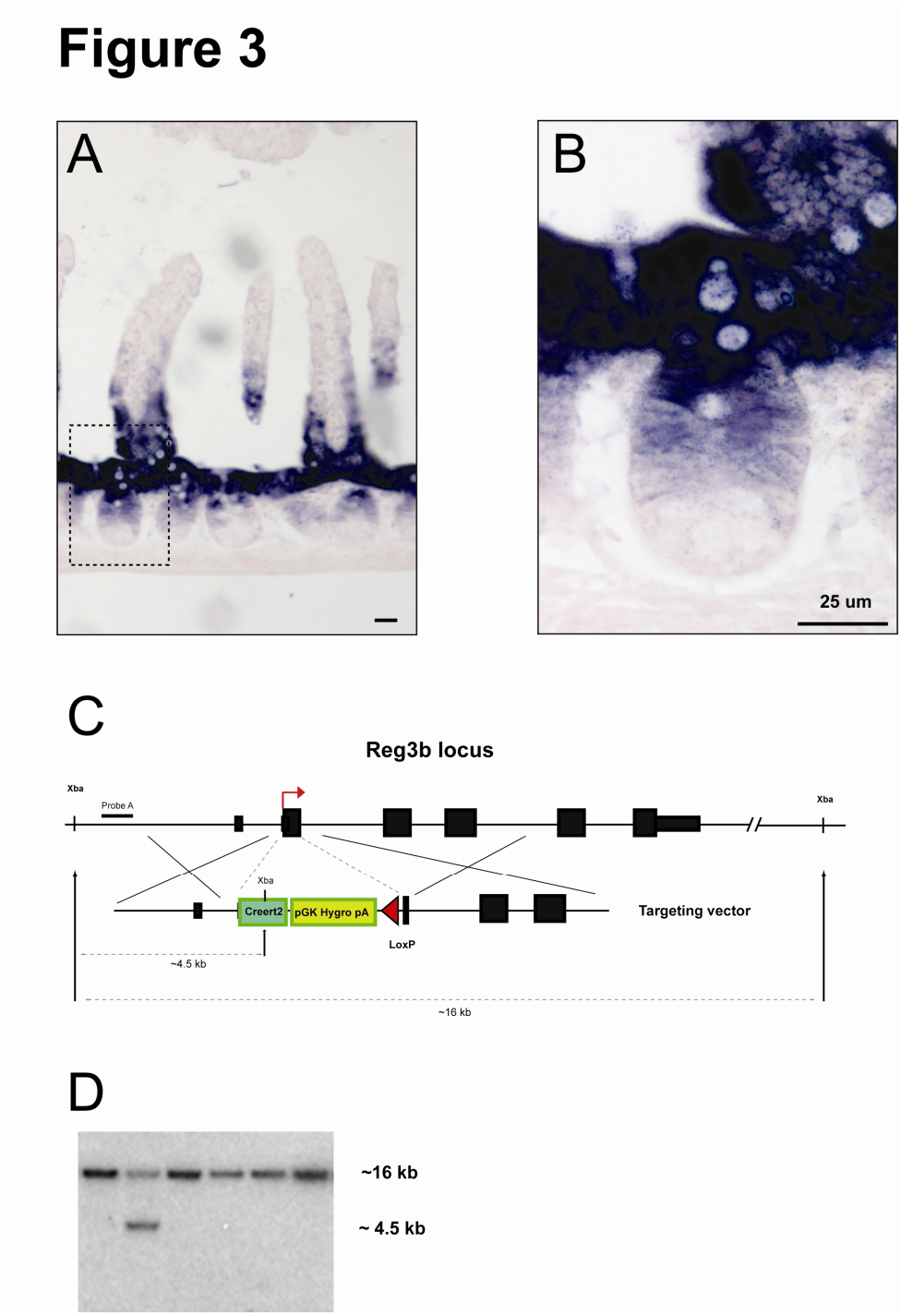


Figure 3 ***Reg3b-Creert2*** **knock-in allele**

a, b, In situ hybridization performed on small intestines of an *Apc*^{min} mouse, illustrating the expression of *Regb* at the upper part of the crypts and at the crypt-villus junctions. Size bars represent 25 µm. **c**, Shows the targeting strategy for the generation of *Reg3b-Creert2* knock-in mice. Hygro, hygromycin resistance cassette. **d**, Southern blot analysis to screen for clones with the correct homologous recombination. ES cell DNA was digested with Xba and hybridized with the probeA as indicated in figure 3c.

Reg3b (also known as Pap) is one of the two flanking genes in the *Reg* cluster. Higher magnification of the *in situ* hybridization experiments shows that the expression of *Reg3b* occurs in the upper part of the crypts and along the crypt-villus junction. Paneth cells and stem cells at crypt bottoms are negative (Figure 3A and B). Figure 3C shows the targeting strategy for the *Reg3b* gene. We replaced most of exon 2 with a tamoxifen inducible Cre variant that we put in frame at the ATG of the gene. A hygromycin^r cassette was included for ES cell selection. One loxP site was introduced for possible future Cre mediated removal of the Reg cluster (see below). We predict that the targeted allele is a functional null, since the signal peptide of the *Reg3b* gene is removed.

Figure 3D shows a southern blot of targeted ES cell clones. One clone harbors the correct homologous recombination. From this ES cell clone, we generated mice. To test the functionality of the targeted allele, we crossed the *Reg3b-Creert2* knock-in mice with a Cre-activatable *Rosa26-LacZ* reporter strain [9] (Figure 4A). Injection of tamoxifen activates the Creert2 fusion enzyme in Reg3b expressing cells. Cre-mediated excision of the roadblock sequence in the *Rosa26-LacZ* reporter should irreversibly mark Reg3b+ cells. The activated LacZ will also be expressed by progeny of these cells since it is expressed from the ubiquitously active *Rosa26* promoter, allowing lineage tracing experiments.

In double heterozygous mice we induced the Cre enzyme with a single tamoxifen injection. Various organs were isolated 1, 2 and 15 days after activation of the Cre enzyme. The one and two day experiments gave almost identical results. We found single LacZ positive cells scattered along the villus in the duodenum (Figure 4B) and ileum (Figure 4D). In the jejunum the number of observed LacZ positive cells was much higher (Figure 4C). The colon was completely negative (Figure 4E). In the 15 day post induction experiment, LacZ staining in the small intestine was lost, underscoring that *Reg3b* is only expressed in transit amplifying cells in the upper part of the crypts and not expressed by the long living stem cells of the intestine. As an example, the jejunum of a 15 day post induction animal is shown in figure 4F. LacZ-positive acinar cells in the pancreas were observed both in the short and the 15 day post-induction time point experiments. Because the cell renewal in the pancreas is much slower than in the intestine, LacZ positive cells can still be observed in the 15 day post induction experiment (Figure 4G).

Non-induced control animals did not show any LacZ staining (data not shown). We also did not observe any LacZ staining in the kidney, skin, liver, stomach and thymus of induced animals at any of the time points (data not shown). From the lineage tracing experiments we conclude that the generated *Reg3b-Creert2* knock-in allele follows endogenous *Reg3b* expression. The generated allele can be used to specifically delete floxed genes in the upper part of the intestinal crypts without recombination in intestinal stem cells.

Figure 4

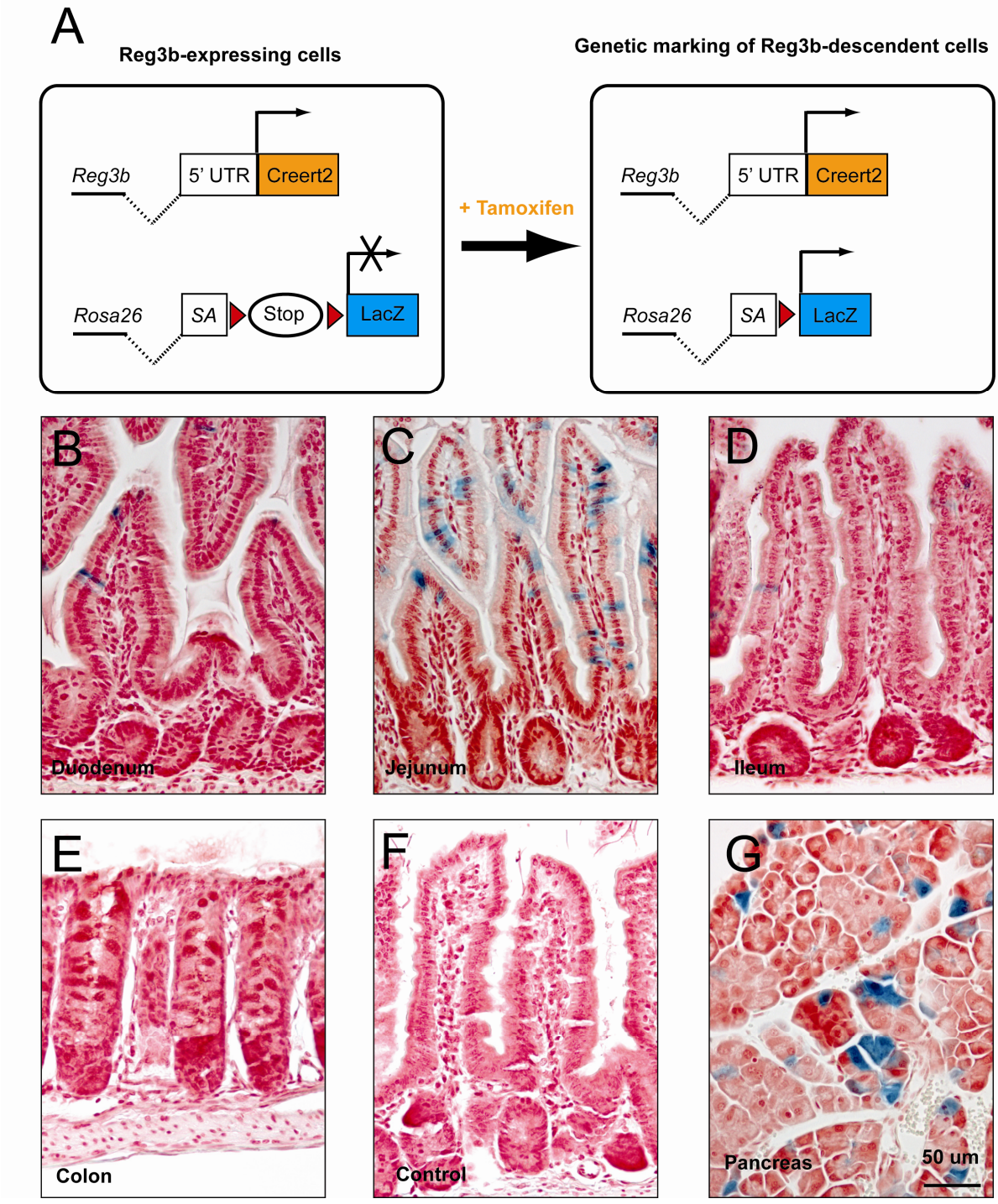


Figure 4 Lineage tracing of Reg3b positive cells in the small intestine and pancreas
a, To test the functionality of the targeted allele we crossed the *Reg3b-Creert2* knock-in mice with a Cre-activatable *Rosa26-LacZ* reporter strain.
b, c, d, 48 hr after tamoxifen injection single LacZ positive cells were found scattered along the villus in the duodenum, jejunum and ileum. The frequency of cells in the jejunum was much higher than in the duodenum and ilium.
e, The colon of these mice did not show any LacZ positive cells.
f, 15 days post induction the LacZ stain was lost in the intestinal epithelium, a picture of the jejunum is shown.
g, In the pancreas LacZ positive acinar cells were still observed 15 days post induction. Size bar represent 50 µm.

Conditional *Reg* cluster locus

To be able to delete the complete *Reg* cluster, we introduced a second loxP site at the other end of the *Reg* cluster. The *Reg3b* ES cells (described above, containing the first loxP site) were targeted with a second construct introducing a loxP site in the *Reg3g* gene. Figure 5A and B shows higher magnifications of *Reg3g* in situ hybridization experiments. Reg3g is expressed in the upper part of the crypts and along the crypt-villus junction. Figure 5C shows the targeting strategy for the *Reg3g* gene. We replaced most of exon 2 with a neomycin^r selection cassette which is flanked by loxP sites, and oriented in the same direction as the loxP site in the *Reg3b* construct. We also introduced a pGK-TK cassette for eventual *in vitro* counter selection. We expect that the targeted allele is a functional null since the leader peptide of the Reg3g protein is removed.

Figure 5D shows a southern blot of targeted ES cell clones, containing two clones with the correct homologous recombination. The two targeting events could have occurred on the same chromosome (Cis) or on opposite chromosomes (Trans) (Figure 5E). To test which ES cells clones were in Cis, we transiently transfected doubly-targeted ES cell clones with a Cre expression plasmid. After two days, DNA was isolated and analyzed by PCR with primers that are located outside the two loxP sites. Figure 5F shows that DNA from three of the five tested clones generated a PCR product. We interpreted these data that these three clones harbored integrations in Cis, while the other two clones harbor the integrations in Trans. Chimeric mice were generated from both types of clones.

Figure 5

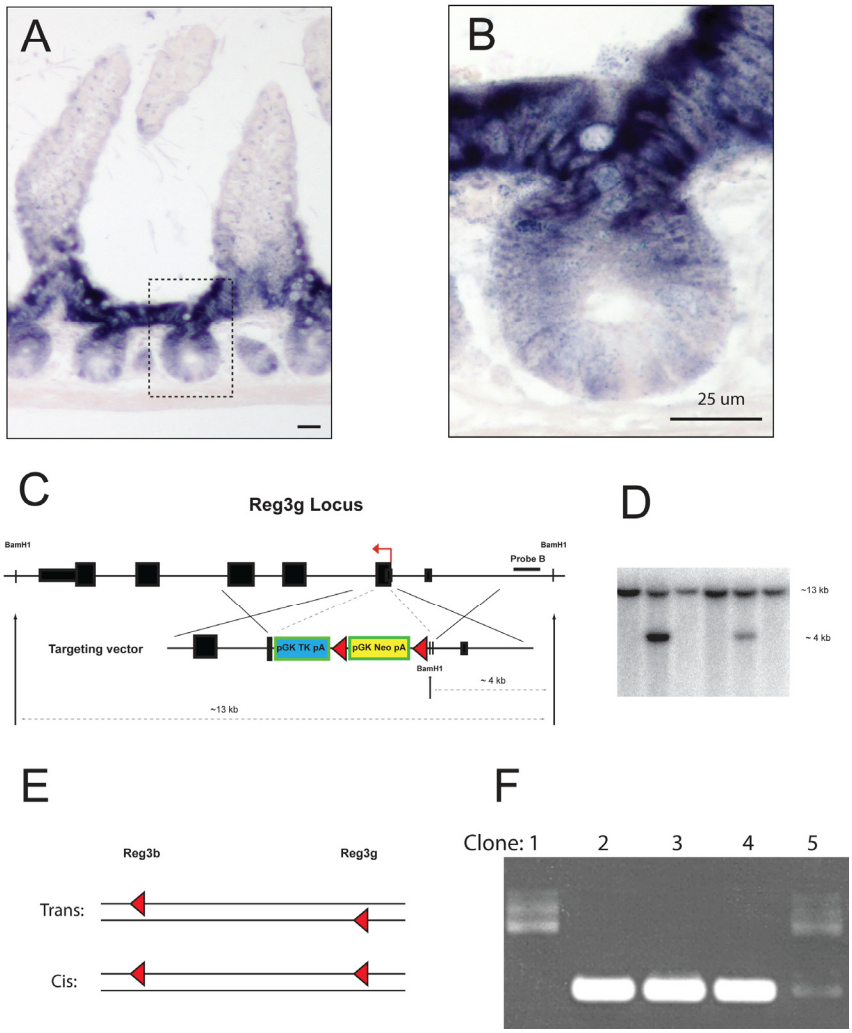


Figure 5 Generation of conditional *Reg* cluster locus.

a, b, In situ hybridization performed on small intestines of an *Apc^{min}* mouse, illustrating the expression of *Reg3g* at the upper parts of the crypts and at the crypt-villus junctions. Size bars represent 25 µm.

c, Shows the targeting strategy for the generation of the *Reg3g* knock-out mice. Neo, neomycin resistance cassette.TK, thymidine kinase.

d, Southern blot analysis to screen for clones with the correct homologous recombination. ES cell DNA was digested with BamH1 and hybridized with the probeB as indicated in figure 5c.

e, Double targeted ES cells can be either in Cis or in Trans.

f, Cis integrated ES cell clones can be distinguished from Trans integrated clones through the formation of a specific PCR product after transient Cre deletion.

Conditional *Reg4* allele

To study the function of all Reg proteins, we also generated a floxed allele for the separate *Reg4* gene located on mouse chromosome 3. The expression pattern of *Reg4* is shown by *in situ* hybridization. In the small intestine, single cells at crypt bottoms are positive as are occasional cells along the villus (Figure 6A). *Reg4* is also expressed at crypt bottoms of the colon (Figure 6B). Figure 6C shows the targeting strategy for the *Reg4* gene. We flanked the second exon of the *Reg4* gene with loxP sites. We expect that Cre mediated deletion will result in a complete null allele because this exon contains the signal peptide of the Reg4 protein. Figure 6D shows a southern blot of targeted ES cell clones containing three clones with the correct homologous recombination. These clones were used for the generation of chimeric mice. The neomycin^{r} selection cassette was flanked by Frt recombination sites and excised *in vivo* by crossing the mice with the FLPeR deleter stain [10].

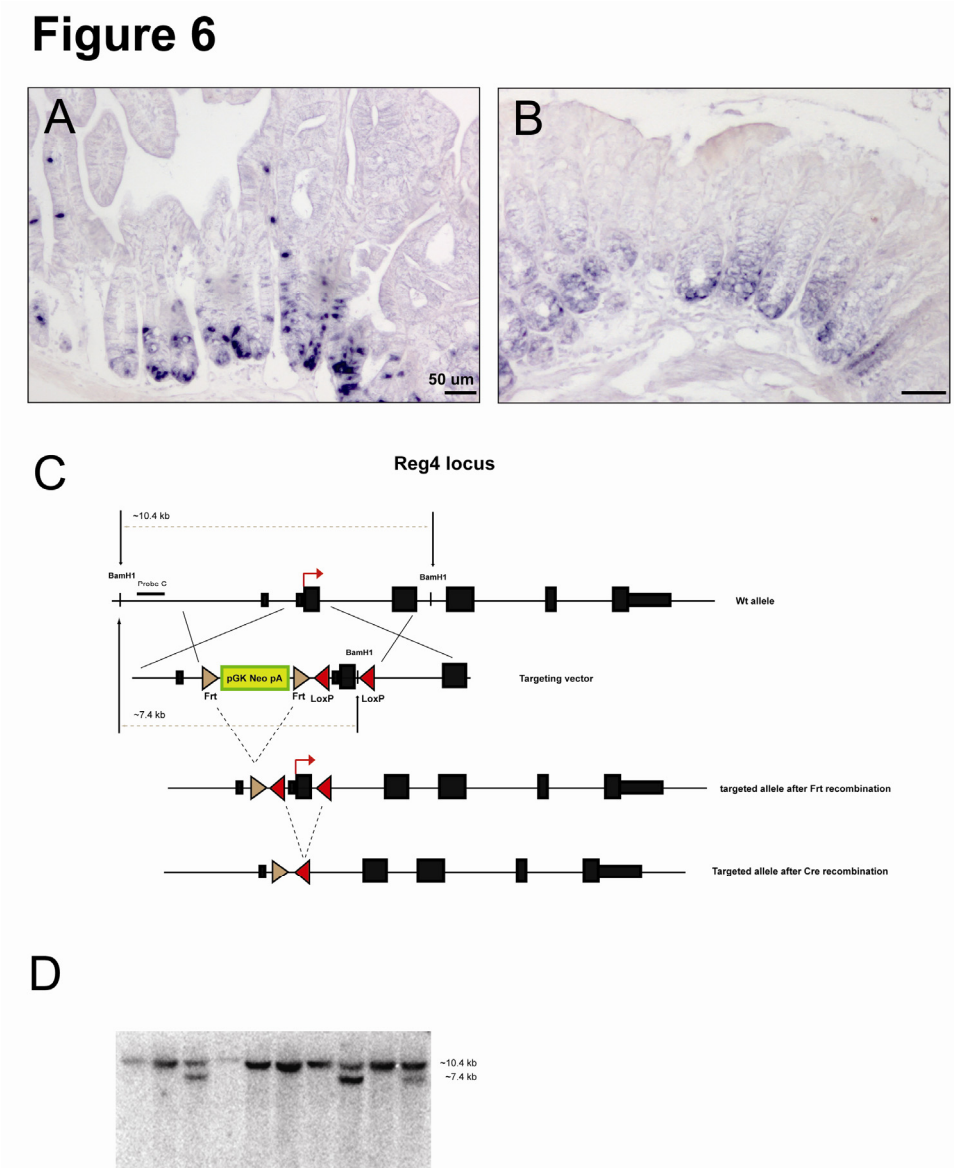


Figure 6 Conditional *Reg4* allele.

a, b, In situ hybridization performed on small intestines of an *Apc^{min}* mouse, illustrating the expression of *Reg4* at single cells in crypts and occasionally along the villus in the small intestine (a) and at crypt bottoms of the colon (b). Size bars represent 50 µm.

c, Shows the targeting strategy for the generation of the floxed *Reg4* mice. Neo, neomycin resistance cassette.

d, Southern blot analysis to screen for clones with the correct homologous recombination. ES cell DNA was digested with BamH1 and hybridized with the probeC as indicated in figure 6c.

Discussion

Reg genes are predominantly expressed by pancreas and intestinal epithelium. In pathological situations these genes are, however, frequently up-regulated in other tissues. Although the Reg proteins are very homologues, we observed that the expression patterns of the various family members are different from each other in wild type intestinal epithelium and pancreas. In intestinal adenomas, most Reg family members are up regulated [3]. *Reg* genes are however not directly regulated by the Wnt pathway [11, 12] but appear to be dependent on different signal(s).

Originally, *Reg1* was identified as being differential expressed in a regenerating islet-screen[2]. Ever since, special attention has been paid to the role of Reg proteins as growth factors. For Reg1, growth-promoting activity has been claimed in the intestine. *Reg1*$^{-/-}$ animals show a decrease in the number of proliferative cells and have a slower cellular migration [13]. The observed effects are however very mild and not consistent with the expression pattern as shown for Reg1 in the paper. Moreover, most other Reg proteins are also expressed in the intestine.

Reg proteins essentially consists of a lectin-binding domain making it unlikely that these proteins act as growth factors. Since Reg expression often reflects tissue injury it is more likely that these proteins serve as acute phase proteins that respond to inflammation or other tissue stress. Recently, it was shown that Reg3g acts as an antimicrobial protein through its lectin domain [14]. It appears very likely that the other family members also have a function in lectin-mediated innate immunity.

To address the role of the *Reg* gene family in the intestine, we are currently cross-breeding the floxed *Reg* cluster and floxed *Reg4* animals with the *Ah-Cre* mouse [6]. In these mice, expression of the *Cre* recombinase is driven by the promoter of the *Cyp1a* gene, which is inducible in a number of tissues upon injection of lipophilic xenobiotic agents such as ß-Naphtoflavone. Analysis of these animals we will likely obtain insight into the role of the complete *Reg* gene family in the intestine.

Material and Methods

Sequence comparison

Amino acid sequences for mouse and human Reg gene homologues and orthologs were downloaded from Ensembl database, release 49 (http://www.ensembl.org). Sequences were aligned using ClustalW v2. [15]. Phylogenetic tree was build with MEGA4 package [16] with Neighbor-Joining method using Poisson-corrected distances. Bootstrap test (1000 replicates) was used to calculate branch support values.

In situ hybridizations

Mouse orthologs of all different *Reg* transcripts were obtained as ESTs from the IMAGE consortium or RZPD. These clones were used for *in vitro* transcription reactions to generate probes for *in situ* hybridizations. Protocols for *in vitro* transcription and *in situ* hybridizations are described elsewhere [17].

Generation of *Reg3b* knock-in, conditional *Reg* cluster and conditional *Reg4* mice

The Reg3b and Reg3g constructs were generated by cloning the various components into the polylinker of pBluescript SK+ (Stratagene) as depicted in figure 3C and figure 5C. The conditional Reg4 construct was generated by cloning the various components into the pl451 vector as depicted in figure 6C. The flanking arms for the constructs were generated by high-fidelity PCR reactions from male 129/Ola-derived IB10 ES cell DNA. All components were sequence-verified. For ES cell targeting the constructs (100 µg) were linearized and transfected into male 129/Ola-derived IB10 ES cells by electroporation (800 V, 3 µF). Recombinant ES cell clones expressing the hygromycin/neomycin gene were selected in medium supplemented with Hygromycin/G418 over a period of 8 days. TK was used as a counter selection for the *Reg3b* and *Reg4* targetings. An ES cell clone harboring the correct homolgous recombination for *Reg3b* gene was targeted a second time with the *Reg3g* construct. Recombinant ES cell clones were screened for the presence of homologous recombinants by Southern blotting. DNA was digested with Xba for *Reg3b* targeting and with BamH1 for the *Reg3g* and *Reg4* targetings and hybridized with probes outside the

constructs. ES cell clones harboring the correct homologous recombination(s) were injected into C57BL/6 blastocysts using standard procedures. In the conditional *Reg4* animals the neomycin expression cassette was flanked by Frt sites and excised *in vivo* by crossing the mice with FlpER deleter strain as previously described [10].

References

1. Zhang, Y. W., Ding, L. S. & Lai, M. D. Reg gene family and human diseases. World J Gastroenterol 9, 2635-41 (2003).
2. Terazono, K. et al. A novel gene activated in regenerating islets. J Biol Chem 263, 2111-4 (1988).
3. Sabates-Bellver, J. et al. Transcriptome profile of human colorectal adenomas. Mol Cancer Res 5, 1263-75 (2007).
4. Lieu, H. T. et al. Reg2 inactivation increases sensitivity to Fas hepatotoxicity and delays liver regeneration post-hepatectomy in mice. Hepatology 44, 1452-64 (2006).
5. Unno, M. et al. Production and characterization of Reg knockout mice: reduced proliferation of pancreatic beta-cells in Reg knockout mice. Diabetes 51 Suppl 3, S478-83 (2002).
6. Ireland, H. et al. Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of beta-catenin. Gastroenterology 126, 1236-46 (2004).
7. Abe, M. et al. Identification of a novel Reg family gene, Reg IIIdelta, and mapping of all three types of Reg family gene in a 75 kilobase mouse genomic region. Gene 246, 111-22 (2000).
8. Miyashita, H. et al. Human REG family genes are tandemly ordered in a 95-kilobase region of chromosome 2p12. FEBS Lett 377, 429-33 (1995).
9. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet 21, 70-1 (1999).
10. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. Widespread recombinase expression using FLPeR (flipper) mice. Genesis 28, 106-10 (2000).
11. Hatzis, P. et al. Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. Mol Cell Biol 28, 2732-44 (2008).
12. Van der Flier, L. G. et al. The Intestinal Wnt/TCF Signature. Gastroenterology 132, 628-32 (2007).
13. Ose, T. et al. Reg I-knockout mice reveal its role in regulation of cell growth that is required in generation and maintenance of the villous structure of small intestine. Oncogene 26, 349-59 (2007).
14. Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science 313, 1126-30 (2006).
15. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-8 (2007).
16. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596-9 (2007).
17. Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 129, 626-38 (2005).

Chapter 8

Summarizing Discussion

Summarizing Discussion

The research described in this thesis primarily focuses on the intestinal epithelium. This is a specialized simple epithelium that lines the gut and performs primary functions of digestion, absorption and forms a barrier against luminal pathogens. The intestinal epithelium is organized in invaginations called crypts and finger-like protrusions called villi. The crypts harbor proliferating stem cells and their transit amplifying daughter cells, while villi are covered with terminally differentiated cell types with specialized functions. There is a constant flow of cells from the bottom of the crypts to the tips of the villi, where the cells are shed into the gut lumen. The intestinal epithelium is almost completely renewed every 3-5 days [1]. Likely because of the high proliferation rate, the intestine is highly sensitive to cancer development. In the Netherlands, every year about 10,000 people are diagnosed to have colorectal cancer, and more than 4000 people die due to this disease (http://www.rivm.nl/vtv/object_document/o1157n17273.html). Malignant transformation of intestinal epithelium is almost invariably initiated by activating Wnt pathway mutations [2, 3]. These mutations either remove the tumor suppressors APC or axin, or activate the proto-oncogene β-catenin. As a common result, β-catenin accumulates in the nucleus, and constitutively binds to the transcription factor TCF4, resulting in transcriptional activation of Wnt/TCF4 target genes. This in turn initiates transformation of intestinal epithelial cells. Physiologically, the Wnt pathway is essential for the maintenance of crypt progenitor proliferation. This was evidenced by mice lacking an intestine-specific member of the Tcf transcription factor family, Tcf4*/Tcf7l2,* that fail to maintain proliferative crypts during late gestation [4]. Moreover, conditional deletion of ß-catenin [5, 6] as well as transgenic expressing of the secreted Dickkopf-1 Wnt inhibitor [7, 8] leads to disappearance of proliferative crypts in adult mice. Because of the intimate connection between Wnt signaling and intestinal biology, we have attempted to unravel the TCF4 target gene program activated by this pathway in crypts and colorectal tumors.

In the chapters 2 and 3, we describe the identification of intestinal Wnt/TCF target genes. We performed expression profiling studies of colorectal cancer cell lines carrying an inducible block of the Wnt signaling pathway. The differentially expressed, β-catenin/TCF responsive, genes were compared with genes up-regulated in prospectively collected colorectal adenomas and/or carcinomas compared with normal mucosa from the same individuals. This resulted in the "intestinal Wnt signature". Through *in situ* hybridization experiments, we showed that Wnt target genes are invariably expressed in adenomas yet could be subdivided into 3 expression modules based on expression in distinct crypt compartments: transit amplifying cells, Paneth cells or stem cells. One of the genes classified as stem cell-specific is the basic helix-loop-helix (bHLH) transcription factor Ascl2 (see below).

The Wnt target genes identified through this approach can be either direct regulated by the TCF/ β-catenin complex or through intermediate transcription factor(s). In chapter 4, we identify direct Wnt targets based on chromatin immunoprecipitation (ChIP)-coupled DNA microarray analysis (ChIP-on-chip). We immunoprecipitated chromatin-bound TCF4 from LS174T CRC cells, and identified the bound DNA sequences through hybridization on DNA microarrays. Testing of these TCF4-bound regions in luciferase-based reporter gene assays demonstrated that these DNA sequences often behave as Wnt-controlled enhancers or promoters.

The extremely high rate of cell renewal makes the intestinal epithelium an attractive organ to study stem cell biology. Recently, we described the first definitive adult intestinal stem cell marker gene: *Lgr5*/Gpr49. *Lgr5* encodes for an orphan G protein–coupled receptor that we originally identified in Wnt target gene expression arrays [9, 10]. Lgr5 expression in the intestine is restricted to Crypt Base Columnar (CBC) cells. Lineage tracing experiments using an *Lgr5-EGFP-ires-CreERT2* knock-in allele crossed with the *Rosa26RLacZ* reporter proved that Lgr5 marks long-lived, pluripotent intestinal stem cells in both the small intestine and the colon [11]. In chapter 5, we have determined a gene signature for these so called CBC cells, isolated on the basis of Lgr5-GFP expression from adult intestines. One of the genes of the stem cell signature was the Wnt target

gene *Ascl2*. The *Ascl2* gene encodes a bHLH transcription factor with an unusually restricted expression pattern, *i.e.* its expression is predominantly detected in extraembryonic tissues [12] and in intestinal epithelium. Transgenic expression of the Ascl2 transcription factor throughout the intestinal epithelium induces crypt hyperplasia and *de novo* crypt formation on villi. Induced deletion of the *Ascl2* gene in adult small intestine leads to disappearance of CBC stem cells within days. The combined results from these gain- and loss-of-function experiments imply that Ascl2 plays an essential role in the maintenance of adult intestinal stem cells.

As mentioned before the intestinal epithelium is a very attractive tissue to study mammalian stem cell biology. In chapter 6 we describe the generation of two new mouse models that will be used for further characterization of intestinal stem cells. The first model is a Cherry-Δtk fusion knocked into the intestinal stem cell specific *Olfm4* gene. In this mouse line, it will be possible to visualize stem cells based on fluorescent Cherry expression and inducibly deplete these cells by activating the Δtk enzyme through ganciclovir injections. The second model is a Cre-activatable multi-color reporter mouse. Upon Cre mediated recombination one out of four different fluorescent colors becomes expressed. Since the color choice is random, neighboring cells will likely be marked with distinct colors, enabling visualization of separate marked clones. We will breed this strain with *Lgr5-GFP-ires-CreERT2* and floxed *APC* mice. Lineage tracing experiments after tamoxifen-induced *Cre* activation in these mice will give us the opportunity to study the dynamics of (cancer) stem cells in the intestine.

In chapter 7, we describe the generation of mouse models that will be used to study the *Regenerating* (*Reg*) gene family. This family encompasses a group of highly homologous, secreted proteins containing a lectin-binding domain. Physiologically, these genes are predominantly expressed in pancreas and in the intestinal epithelium. Under pathological conditions, *Reg* genes are however, frequently up regulated in diverse tissues. The exact function of the Reg proteins is unclear. In order to study the function of the entire *Reg* gene family in the intestine, we took a genetic loss-of-function approach in mice. Both the genes from the *Reg* cluster and the *Reg4* gene were flanked by loxP sites facilitating Cre mediated recombination. To study the role of the Reg gene family in the intestine, we will cross-breed the floxed *Reg* cluster and floxed *Reg4* animals with the *Ah-Cre* mouse [6]. In these mice, expression of *Cre* recombinase is driven by the promoter of the *Cyp1a* gene, which is inducible in a number of tissues upon injection of lipophilic xenobiotic agents such as ß-Naphtoflavone.

References

1. Barker, N., van de Wetering, M. & Clevers, H. The intestinal stem cell. Genes Dev 22, 1856-64 (2008).
2. Korinek, V. et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 275, 1784-7 (1997).
3. Morin, P. J. et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 275, 1787-90 (1997).
4. Korinek, V. et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 19, 379-83 (1998).
5. 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 27, 7551-9 (2007).
6. Ireland, H. et al. Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of beta-catenin. Gastroenterology 126, 1236-46 (2004).
7. Kuhnert, F. et al. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A 101, 266-71 (2004).

8. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev 17, 1709-13 (2003).
9. van de Wetering, M. et al. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 111, 241-50 (2002).
10. Van der Flier, L. G. et al. The Intestinal Wnt/TCF Signature. Gastroenterology 132, 628-32 (2007).
11. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449, 1003-7 (2007).
12. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J. & Joyner, A. L. Essential role of Mash-2 in extraembryonic development. Nature 371, 333-6 (1994).

Samenvatting

Samenvatting

Het onderzoek beschreven in dit proefschrift gaat voornamelijk over het darmepitheel. Dit is de laag cellen die de binnenkant van de darm bekleedt. Deze laag cellen is van groot belang voor de spijsvertering en vormt bovendien een barrière tegen ziekteverwekkende micro-organismen. Het darmepitheel heeft een groot oppervlakte doordat het bestaat uit instulpingen (crypten) en uitstulpingen (darmvlokken). De darmvlokken zijn bekleed met rijpe gespecialiseerde cellen zoals de slijmproducerende slijmbekercellen, hormoon uitscheidende enteroendocrine cellen en enterocyten die voedingsstoffen opnemen. Deze gespecialiseerde cellen kunnen zelf niet meer delen maar stammen af van actief delende stamcellen die zich bevinden in de crypten. Naast stamcellen bevinden zich in de crypten ook "transitamplifying" cellen en Paneth cellen. Transitamplifying cellen zijn de dochtercellen van stamcellen die nog een paar keer kunnen delen voor ze volledig uitrijpen. Paneth cellen zijn het vierde gedifferentieerde cel type van de darm, en vormen een uitzondering doordat ze niet omhoog migreren naar de darmvlokken maar in de crypten blijven. Deze cellen spelen een rol bij de afweer tegen ziekteverwekkende micro-organismen. Het darmepitheel wordt continu vernieuwd. In 3 tot 5 dagen migreren en differentiëren de cellen vanuit de crypt naar de tip van de darmvlokken, waar de cellen uiteindelijk loslaten en worden afgevoerd via het darmkanaal. Het darmepitheel is het snelste zelfvernieuwend weefsel in volwassen zoogdieren. Eén van de meest belangrijke signaleringsroutes om dit proces van celdeling goed te laten verlopen is de zogenaamde 'Wnt route'. Deze signaleringsroute reguleert de interactie tussen het eiwit β-catenine en TCF transcriptie factoren. Deze twee eiwitten vormen samen een DNA-bindend complex dat de expressie van andere genen reguleert.

Waarschijnlijk maakt de grote hoeveelheid celdeling in de darm dit orgaan ook zeer gevoelig voor de ontwikkeling van kanker. In Nederland wordt bij ongeveer 10000 mensen per jaar de diagnose darmkanker vastgesteld en ongeveer 4000 mensen per jaar overlijden aan de gevolgen van deze ziekte. Darmkanker wordt in bijna alle gevallen veroorzaakt door activerende mutaties (veranderingen) in de Wnt signaleringsroute. Ten gevolge van deze mutaties hoopt het eiwit β-catenine zich op in de celkern en bindt het aan de transcriptie factor TCF. Dit resulteert in ongecontroleerde celdeling en is de eerste stap in de ontwikkeling van darmkanker.

Omdat de Wnt signaleringsroute zo'n essentiële rol speelt in zowel de fysiologie van de gezonde darm als bij darmkanker hebben we geprobeerd de genen die door Wnt/TCF gereguleerd worden (target genen) te identificeren. In de hoofdstukken 2 en 3 van dit proefschrift beschrijven we de expressie van genen in darmkanker cellijnen waar we de Wnt signaleringsroute induceerbaar hebben geremd. De door Wnt/TCF gereguleerde genen zijn vervolgens vergeleken met RNA expressie profielen van colorectale adenoma en adenocarcinoma samples van patiënten. Dit resulteerde in een set van Wnt target genen in de darm. De precieze expressie patronen van deze set genen is vervolgens, door middel van zogenaamde *in situ* hybridisatie experimenten, bestudeerd in een muizendarmkanker model. Uit deze experimenten bleek dat Wnt target genen altijd tot expressie komen in adenomen maar dat de expressie in normale crypten wordt gevonden in Paneth cellen, transitamplifying cellen of stamcellen. Binnen de groep stamcel specifieke Wnt target genen vonden we ondermeer de transcriptiefactor ASCL2 (zie hieronder).

De target genen die we in hoofdstuk 2 en 3 geïdentificeerd hebben kunnen zowel direct als indirect, door middel van tussenliggende transcriptiefactoren, onder invloed van de Wnt signaleringsroute staan. In hoofdstuk 4 hebben we directe Wnt targets geïdentificeerd door te kijken waar TCF4 precies aan het DNA bindt. De TCF4 gebonden regio's in het DNA blijken vaak rondom "expressie" target genen te liggen en gedragen zich vaak als Wnt gecontroleerde enhancers of promotors (regel regio's).

De extreem hoge celvervanging en het duidelijke functionele en histologische onderscheid tussen crypt en darmvlok cellen maken het darmepitheel een erg aantrekkelijk orgaan voor het bestuderen van stamcellen en hun nakomelingen. De darmen van de mens en de muis lijken sterk op elkaar.

Genetische manipulatie van de muis stelt ons in staat om de directe functie van genen te bestuderen. Recent hebben we het eerste definitieve volwassen darmstamcel marker gen beschreven: Lgr5/Gpr49. In genetische markeringsstudies hebben we laten zien dat Lgr5/Gpr49 alleen specifiek door darmstamcellen gemaakt wordt. In hoofdstuk 5 hebben we deze Lgr5 positieve stamcellen geïsoleerd, waardoor we in staat waren het RNA expressie patroon van deze cellen te bepalen. Eén van de stamcel specifieke genen bleek inderdaad Ascl2 te zijn. Om de functie van dit gen in de darm te bestuderen hebben we Ascl2 tot over-expressie gebracht in de muizen darm. Dit resulteerde in meer celdeling en nieuwe crypt vorming in de darmvlokken. Het induceerbare uitschakelen van het Ascl2 gen in de darm resulteerde in het verdwijnen van stamcellen van de darm. De combinatie van deze genetische proeven toont aan dat Ascl2 een essentiële rol speelt in het handhaven van de volwassen darmstamcellen.

In hoofdstuk 6 worden twee muismodellen beschreven die gegenereerd zijn voor het verder karakteriseren van darmstamcellen. In het eerste model is een stamcelspecifiek gen vervangen door een fluorescerend eiwit dat isolatie en visualisatie van deze cellen mogelijk maakt. Dit fluorescente eiwit wordt gevolgd door een speciaal enzym, dat normaal niet tot expressie komt. Dit enzym veroorzaakt celdood en kan geactiveerd worden door toevoeging van een specifieke stof. Door het enzym te activeren gaan darmstamcellen dood, waardoor we de darm in afwezigheid van deze cellen kunnen bestuderen. Het tweede model is een induceerbare multi-kleuren reporter. In deze muis is het mogelijk om onafhankelijke cellen te markeren met één van vier fluorescente kleuren. We willen dit klonale markeringsmodel gaan gebruiken om de dynamiek van (kanker)stamcellen van de darm te bestuderen.

In hoofdstuk 7 beschrijven we het genereren van verschillende muis modellen die in de toekomst kunnen worden gebruikt voor het bestuderen van de Regenerating (Reg) genen familie in de darm. Deze Reg genen familie omvat een groep van zeer homologe uitgescheiden eiwitten. Fysiologisch worden deze eiwitten vooral gemaakt in de darm en de pancreas. Onder pathologische condities, o.a. darmkanker, komen ze echter in diverse weefsels tot expressie. De precieze functie van deze groep eiwitten is niet bekend. Met behulp van deze induceerbare knock-out muizen waarbij we de Reg familie genen kunnen uitschakelen zijn we nu in staat de functie van deze eiwitten te bestuderen.

Dankwoord

Na mijn studie ben ik bij de onderzoeksafdeling van Numico gaan werken, waar het plan om te promoveren al vrij snel bij mij ontstond. Het duurde echter even, voor het tot me doordrong dat ik dit ook gewoon in een onderzoeksinstituut als OIO zou kunnen doen. Toen die gedachte eenmaal was geland, ben ik meteen ook maar voor de Moleculaire Biologie gegaan. Niet direct een logische keuze, gezien het feit dat ik tijdens mijn studie alle vakken in die richting zoveel mogelijk heb proberen te vermijden. De onderzoeksgroep van Hans Clevers kende ik uit de wetenschapsbijlage van de krant. Na een e-mail, twee gesprekken, en een telefoontje kon ik zowaar beginnen!

Hans, allereerst wil ik je bedanken voor het feit dat je mensen durft aan te nemen die niet eens weten wat een restrictie-enzym is. Daarnaast wil ik je ook ontzettend bedanken voor de geweldige tijd die ik in het lab heb gehad. Ik heb de afgelopen vijf jaar met veel plezier gewerkt en ongelofelijk veel van je geleerd. Ik heb erg veel bewondering voor je positieve instelling en je pragmatische manier van werken.

Verder wil ik natuurlijk al mijn (ex-)collega's van in en om het Clevers lab bedanken. Het is gewoon een feit dat ik zonder al jullie input dit boekje nooit vol had kunnen krijgen! In het bijzonder zou ik graag noemen Marc, Hugo en Pantelis. Met jullie heb ik ook buiten het werk veel gedaan. Marc en Hugo: met jullie zou ik graag nóg eens naar Barcelona gaan, om daar in een tapas-restaurant de volledige kaart te bestellen. Pantelis: it is always interesting to see how the Poker rules can change, depending on yóur cards! Anyway, I am now ready to play the Caretta caretta variant!

Buiten het lab wil ik graag bedanken Antoon, Marc en Wouter. Het is geweldig om een aantal goede vrienden te hebben met wie ik werkelijk over alle onderwerpen kan praten en discussiëren, maar die bovenal met de vele grote en kleine stappen in mijn leven meeleven.

Mijn ouders, Wiesje en Henk. Wat is het toch fijn om jullie als familie te hebben. Jullie hebben altijd interesse in wat ik doe, en ik heb het gevoel dat ik altijd op jullie kan terugvallen.

Als laatste wil ik graag noemen Dorine en Justine. Dorine, het is geweldig om samen met je te zijn, en wat fantastisch dat Justine er nu bij is! Ik ben blij met jouw mening, maar geniet vooral ook van je relativeringsvermogen. Werk is leuk, maar er zijn zoveel leuke(re) dingen! Met teksten als "ach, die drie wetenschappers in de wereld die in dat artikeltje van jou geïnteresseerd zijn" wist je me tijdens dit onderzoek met beide benen op de grond te houden. Nu deze promotie is afgerond is er gelukkig weer wat meer tijd voor zaken als zeilen en bergwandelen. Ik hoop dat we de komende jaren ons werk kunnen blijven relativeren, en tijd nemen voor dingen die écht belangrijk zijn!

Laurens

Curriculum vitae

Laurens Gerrit van der Flier is geboren op 11 april 1977 in Leidschendam. In 1996 behaalde hij zijn VWO diploma aan het Thomas a Kempis College in Arnhem. Na het behalen van de propedeuse Biologie aan de Universiteit Utrecht in 1997 rondde hij in 2001 de bovenbouw-studie Fundamentele BioMedische Wetenschappen met succes af aan dezelfde universiteit. Tijdens deze studie liep hij onderzoeksstages bij de leerstoelgroep Stofwisselingsfysiologie van de Universiteit Utrecht (onder begeleiding van dr. W.J.A. van Marrewijk en Prof. dr. D.J. van der Horst) en bij de afdeling Longziekten van het Speciaal Klinisch Immunolgisch Laboratorium in het Academisch Medisch Centrum te Amsterdam (onder begeleiding van dr. R. Lutter). Vervolgens werkte hij ruim twee jaar als junior onderzoeker bij de Afdeling Biomedical Research van Numico-Research in Wageningen. In januari 2004 startte hij als onderzoeker in opleiding bij de onderzoeksgroep van Prof. dr. H.C. Clevers in het Hubrecht Laboratorium te Utrecht. De resultaten van dit onderzoek zijn in dit proefschrift beschreven.

List of publications

Van der Flier LG, Van Gijn ME, Hatzis P, Kujala P, Haegebarth A, Stange DE, Begthel H, van den Born M, Guryev V, Oving I, Van Es JH, Barker N, Peters PJ, Van de Wetering M and Clevers H. Transcription factor Achaete scute-like 2 (Ascl2) controls intestinal cell fate. Cell, in press.

Fan XY, van den Berg A, Snoek M, van der Flier LG, Smids B, Jansen HM, Liu RY, Lutter R. Low concentrations of arginine enhance inflammatory mediator production by airway epithelial cells. Submitted for publication.

Van der Flier LG, Clevers H. Stem cells, self renewal, and differentiation in the intestinal epithelium. Annual Reviews of Physiology. 2008 Sep 22. [Epub ahead of print].

Hatzis P, van der Flier LG, van Driel MA, Guryev V, Nielsen F, Denissov S, Nijman IJ, Koster J, Santo EE, Welboren W, Versteeg R, Cuppen E, van de Wetering M, Clevers H, Stunnenberg HG. Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. Mol Cell Biol. 2008 Apr;28(8):2732-44.

Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, Laczko E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Gomes V, Pastorelli A, Faggiani R, Anti M, Jiricny J, Clevers H, Marra G. Transcriptome profile of human colorectal adenomas. Mol Cancer Res. 2007 Dec;5(12):1263-75.

Van der Flier LG, Sabates-Bellver J, Oving I, Haegebarth A, De Palo M, Anti M, Van Gijn ME, Suijkerbuijk S, Van de Wetering M, Marra G, Clevers H. The Intestinal Wnt/TCF Signature. Gastroenterology. 2007 Feb;132(2):628-32.

Gorselink M, Vaessen SF, van der Flier LG, Leenders I, Kegler D, Caldenhoven E, van der Beek E, van Helvoort A. Mass-dependent decline of skeletal muscle function in cancer cachexia. Muscle Nerve. 2006 May;33(5):691-3.