# WNT signaling, hypoxia and GATA6: Their role in colorectal cancer, stem cells and disease progression

**Gavin Whissell**

TESI DOCTORAL UPF / 2013

DIRECTOR:

**Dr. Eduard Batlle Gómez:** Institute for Research in Biomedicine
Oncology Program
Colorectal Cancer Laboratory


DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES (CEXS)

# ACKNOWLEDGMENTS

This thesis is the product of seven years of experimental research at the Institute for Research in Biomedicine (IRB) Barcelona as a graduate student at the Universitat Pompeu Fabra (UPF).

I have met many outstanding people who have influenced my life throughout my time as a PhD student in and out of the lab. This academic journey has been filled with many ups and downs, but ultimately has ended well with the product of three manuscripts. I have gained a vast repertoire of technical skills and a thorough understating of complex and divergent scientific topics. Without the support of all those around me this achievement would not have been possible.

First of all, I would like to thank my supervisor Dr. Eduard Batlle for allowing me the opportunity to join his laboratory. Eduard has given me a great deal of freedom and independence to pursue the various projects that I have undertaken and participated in throughout out my stay in his laboratory. His experience and knowledge has helped me grow as scientist and has allowed me explore possibilities I would have never considered. He has also taught me the importance of looking beyond the data in order to construct compelling stories. I thank him for his time, patience, and dedication to science.

I would like to express my gratitude to the IRB Barcelona for selecting and providing me with an IRB Fellowship. The IRB is an outstanding institution and I feel truly privileged to have worked in such a modern and well equipped institution among its extraordinary scientists that place IRB among the best science centers in Spain.

Throughout my graduate studies several colleagues have gone above and beyond to assist me. My switch from the field of Environmental Microbiology to Oncology was made possible thanks to the initial training I received from Elena, for this I am very grateful. Special thanks to Xavi, Elisa M, Peter, Alex and Marta for their time, discussions, advice, friendship and for being there when I needed them most. Many thanks to Mercedes, Elena and Francisco for taking the time to proofread my thesis. I would like to thank the remaining colleagues in the laboratory, past and present, for their continued support throughout the years: Anna, Annie, Carme, Clara, Daniele, Elisa E, Enza, Guiomar, Isabella, Juan Luis, Mar and Sergio.

I would also like to thank Kyra and Jordi, from the Cell and Developmental Biology Program, for allowing me the opportunity to collaborate with them. They are exceptional scientists and I feel privileged to have had to chance to work with them. This fruitful collaboration led to the publication of the manuscript presented in Chapter 3.

The continued support from the core facilities at IRB Barcelona, namely the Advanced Digital Microscopy (Anna and Lidia), Biostatistics / Bioinformatics Unit (Evarist, David and Camille) , Functional Genomics (Herbert and Annie), Histology service (Begoña) and Protein Expression core facilities (Nick) were instrumental in making this PhD a success. Additionally, the Cytometry department (Jauma and Sonia) at Scientific and Technological Centers of the University of Barcelona (CCiTUB) provided unparalleled expertise in all cytometry related experiments.

Very special thanks to Herbert for fruitful discussions, advice and his friendship.

I am thankful to my friends out of the laboratory namely, Juan Cruz, Eli, Horacio, Larisa, Nahuel, Federico, Alina, Facundo and Mariano for great times and memories during my stay in Barcelona.

During my PhD a great deal in my life has changed; I married the love of my life and our two beautiful children, Nina and Logan, were born I would like to thank my family for their continued support throughout my graduate studies. Particularly, I wish to express my deepest gratitude to my wife Mercedes for her, support, infinite patience, comprehension and unconditional love.

# ABSTRACT

The recent discovery that intestinal adenomas are built on a stem cell hierarchy has raised great interest in understanding the signals that maintain tumor stem cells. We unveiled the existence of a transcriptional circuit dedicated to prevent the expansion of adenoma stem cells during the onset of CRC. This circuit was found to be controlled by the transcription factor GATA6, which directly activated the expression of the WNT pathway component LGR5 and repressed BMP levels in adenomas. As a result of this mechanism, two cell compartments were established in adenomas: a BMP positive zone that comprises differentiated tumor cells and a BMP negative niche that hosts adenoma stem cells. Genetic deletion of *Gata6* elevated BMP levels in tumors, which in turn inhibited adenoma stem cell self-renewal and intestinal tumorigenesis. These findings represent a key contribution to understand the mechanisms that regulate the tumor stem cell hierarchy and reveal for first time the existence of a niche that protects AdSCs from BMP signals.

In subsequent studies, we further explore the role of GATA factors in EMT both in Drosophila and mammalian models. During the formation of the Drosophila endoderm relatively static epithelial cells have to convert to highly migratory cells to form a large part of the intestinal tract. We showed that the Drosophila GATA factor Srp is necessary to induce EMT, which resulted in relocalization of E-cadherin protein without altering its expression. Mammalian GATA4 and GATA6 also induced an EMT when ectopically expressed in culture. GATA factor Srp and GATA6 both transcriptionally inhibited Crumbs orthologues, delocalized E-cadherin and activated mesenchymal genes. This work has uncovered a novel evolutionarily conserved alternative route to EMT that is Snail independent. This discovery could have important clinical implications in pathogenesis, namely in cancer progression.

Further studies aimed to characterize WNT signaling during CRC progression. WNT signaling is required for the maintenance of colorectal cancer stem cells (CRC-SCs) but paradoxically low levels of WNT genes associate with a higher risk of disease relapse. We identified a core expression program driven by beta-catenin/TCF in CRC-SCs, which was shown to be a strict indicator of the dependency on WNT signaling for growth. We showed that during disease progression, a subset of WNT genes is downregulated as a consequence of tumor hypoxia. This response characterized a group of patients displaying very high risk of cancer recurrence after therapy. Hypoxia-induced genes were prominently expressed at invasion fronts coinciding with lower expression of the WNT program. Therefore, hypoxia refines the WNT program in CRC-SCs during the acquisition of a malignant phenotype. This finding has important

implications in CRC progression, as it explains how some CRC-SC genes (like EPHB2) are shut down during the progression of the disease and directly challenges recent high impact work in the field that points to methylation as the mechanism responsible for the refinement of WNT signaling. Moreover it identifies hypoxia as a culprit in this invasive cell phenotype. These findings may have important clinical implications in treatment of CRC.

# RESUMEN

El reciente descubrimiento de que los adenomas intestinales mantienen una jerarquía similar a la del epitelio intestinal normal, ha despertado gran interés por entender las señales que especifican y mantienen a las células madre tumorales. En este trabajo, damos a conocer la existencia de un circuito transcripcional dedicado a prevenir la expansión de células madre de adenoma durante el inicio del cáncer colorectal (CCR). Este circuito está controlado por el factor de transcripción GATA6, el cual activa directamente el componente de la vía de señalización WNT LGR5, y reprime los niveles de BMP en adenomas. Como resultado de este mecanismo, se establecen dos compartimentos celulares en los adenomas: una zona BMP positiva que contiene células tumorales diferenciadas y un nicho BMP negativo que aloja células madre de adenoma. La deleción genética de *Gata6* eleva los niveles de BMP en tumores, lo que a su vez inhibe la auto-renovación de células madre de adenoma y la tumorigénesis intestinal. Estos hallazgos representan una contribución fundamental para comprender los mecanismos que regulan la jerarquía de las células madre tumorales y revelan por primera vez la existencia de un nicho que protege a las células madre de adenoma de señales BMP.

En estudios posteriores, continuamos explorando el rol de los factores GATA en la transición epitelio-mesénquima tanto en modelos de Drosophila como de mamífero. Durante el desarrollo del endodermo en Drosophila, células epiteliales relativamente estáticas se convierten en células migratorias para formar una gran parte del tracto intestinal a través del proceso de transición epitelio-mesénquima. Hemos demostrado que el factor GATA -en Drosophila denominado Srp- es necesario para inducir una transición epitelio-mesénquima que da lugar a la relocalización de la proteína E-cadherina sin alterar su expresión. Los factores GATA4 y GATA6 de mamífero también inducen transición epitelio-mesénquima cuando son expresados de forma ectópica en cultivos celulares. Tanto el factor GATA Srp como GATA6 inhiben la transcripción de ortólogos de Crumbs, deslocalizan la E-cadherina y activan genes mesenquimales. Nuestro trabajo ha dado a conocer una nueva ruta alternativa para la transición epitelio-mesénquima que se encuentra conservada evolutivamente y que es independiente de Snail. Este descubrimiento podría tener importantes implicaciones clínicas en patogenia, y más específicamente en la progresión del cáncer.

Un tercer aspecto del trabajo presentado se centra en la caracterización de la vía de señalización WNT durante la progresión del CCR. La señalización WNT es necesaria para el mantenimiento de las células madre de cáncer colorectal (CM-CCR), aunque paradójicamente, bajos niveles de genes WNT se asocian con un mayor riesgo de recaída de la enfermedad. Hemos identificado un programa básico de expresión impulsado por beta-catenina/TCF en CM-CCR, que se ha demostrado ser un indicador estricto de la dependencia de señalización WNT para el crecimiento. Hemos demostrado que durante la progresión de la enfermedad, se suprime la expresión de un subconjunto de

genes WNT como consecuencia de hipoxia tumoral. Esta respuesta caracteriza a un grupo de pacientes con un muy alto riesgo de reincidencia del cáncer después de la terapia. Los genes inducidos por hipoxia se expresan de manera prominente en frentes de invasión coincidiendo con una menor actividad de la vía del programa WNT. Por lo tanto, la hipoxia refina el programa WNT en CM-CCR durante la adquisición de un fenotipo maligno. Este hallazgo tiene implicaciones importantes para la progresión del CCR, porque explica cómo algunos genes de CM-CCR (como receptores EPHB2) se apagan durante la progresión de la enfermedad y directamente desafía trabajos de alto impacto recientemente reportados en el área, que apuestan por la metilación como mecanismo responsable del refinamiento del programa WNT. Por otra parte, identifica a la hipoxia como un culpable de este fenotipo de células invasivas. Estos hallazgos podrían tener importantes implicaciones clínicas en el tratamiento del cáncer colorectal.

# PREFACE

The body of this thesis is composed of three chapters, preceded by an introduction and followed by a discussion, future prospective and conclusion section. Each chapter is based on a scientific article which at the time of thesis submission was either published or under revision. The experimental work was carried out in the Colorectal Cancer laboratory at the Institute for Research in Biomedicine (IRB) located at the Parc Cientific de Barcelona (PCB). It would not have been possible to complete this Thesis without the constant guidance and supervision of Dr. Eduard Batlle and the collaboration of co-authors in all three articles.

In a first chapter (Chapter 2), the work related with the identification of the transcription factor GATA6 as a mediator of a transcriptional circuit that regulates the expansion of adenoma stem cells is presented. GATA6 mediates this through repression of BMP signals as well as repression of WNT pathway inhibitors and activation of the WNT/intestinal stem cell signaling component LGR5. This circuit aids in delineating tumor heterogeneity by establishing more differentiated (BMP positive) and undifferentiated (BMP low) tumor cell zones. This article has recently been submitted for publication to Nature Cell Biology and is currently under revision. The author of this thesis was responsible for the planning, execution and analysis of most experimental work presented herein and contributed by writing an original manuscript draft.

In a second chapter (Chapter 3), the work presented resulted from collaboration with the "Morphogenesis in *Drosophila*" group led by Dr. Casanova at our Institute. In this work, we unveiled a new kind of EMT (endodermal-EMT) that is conserved from flies to mammals, which is activated by GATA factors (Srp and GATA6). This novel EMT is Snail independent and is characterized by relocalization of membranous E-cadherin, yet does not affect its protein levels. This article was published in Developmental Cell, 2011, 21 (6): 1051-1061. Dr. Campbell, first author of this article, and her colleagues performed all the work with *Drosophila* while the author of this thesis carried out the investigation with mammalian cells and was granted second authorship.

In the third chapter (Chapter 4), we investigated avenues by which WNT target genes are modulated throughout CRC progression. We identified hypoxia as a negative regulator of a subset of WNT/CRC-SC genes in aggressive CRC and show that hypoxia induced genes are highly expressed at tumor invasion fronts. This discovery explains how some CRC-SC genes, like EPHB2, are silenced during disease progression without impinging on WNT pathway activity. This article is currently under revision in Cell Reports. The author of this thesis was responsible for the planning, execution and analysis of most experimental work presented herein and contributed by writing an original manuscript draft.

In addition, results from this Thesis were communicated in the following conferences:

“30 Years of Wnt Signaling” Oral presentation by E. Batlle EMBO Conference. Egmond aan Zee, Netherlands. June 27-July 1st, 2012

“From Colon Stem Cells to Colorectal Cancer” Oral presentation by E. Batlle. Gordon Conference on Stem Cells and Cancer. Les Diablerets, Switzerland. April 21-26, 2013.

# Table of Contents

# CHAPTER 1: INTRODUCTION

# 1. THE INTESTINE

## 1.1 Structure and organization of the intestinal tract

The intestinal tract is a tubular structure that extends from stomach to the rectum (Figure 1A). It consists of 4 concentric layers of tissue (mucosa, submucosa, muscularis externa and serosa). The mucosa comprises the epithelium, lamina propria and muscularis mucosae. The folded epithelium of the mucosa, a mono layer of columnar cells, is the inner most portion of the intestinal tube that faces the luminal content of the intestine (Figure 1B). It possesses specialized cells that primarily function in digestion, water and nutrient absorption, secretion, fecal compaction and form a protective barrier against pathogens. The intestine is anatomically and functionally divided into two main parts along the cephalocaudal axis, the small intestine (duodenum, jejunum and ileum) and the large intestine or colon (proximal and distal colon) (Figure 1A). The small intestine is organized into tubular crypts that invaginate into the underlying mesenchyme and finger-like villi that project into the lumen (collectively referred to as crypt-villus units) whereas the colon possesses crypts and no villi (Figure 1A and 2). The villi in the small intestine are paramount in maximizing the intestinal surface area for absorption of vitamins and nutrients. The colon, devoid of villi, serves to absorb water and salts for stool compaction. The crypts and villi (or surface epithelial cells in the colon) constitute both the proliferative and terminally differentiated compartments, respectively. The epithelial content of the flat colonic epithelial surface and small intestinal villi are polyclonal as they are replenished by several neighbouring crypts where stem cells residing at the base of the crypts permanently replenish the arrays of cells that constitute the intestinal epithelium.

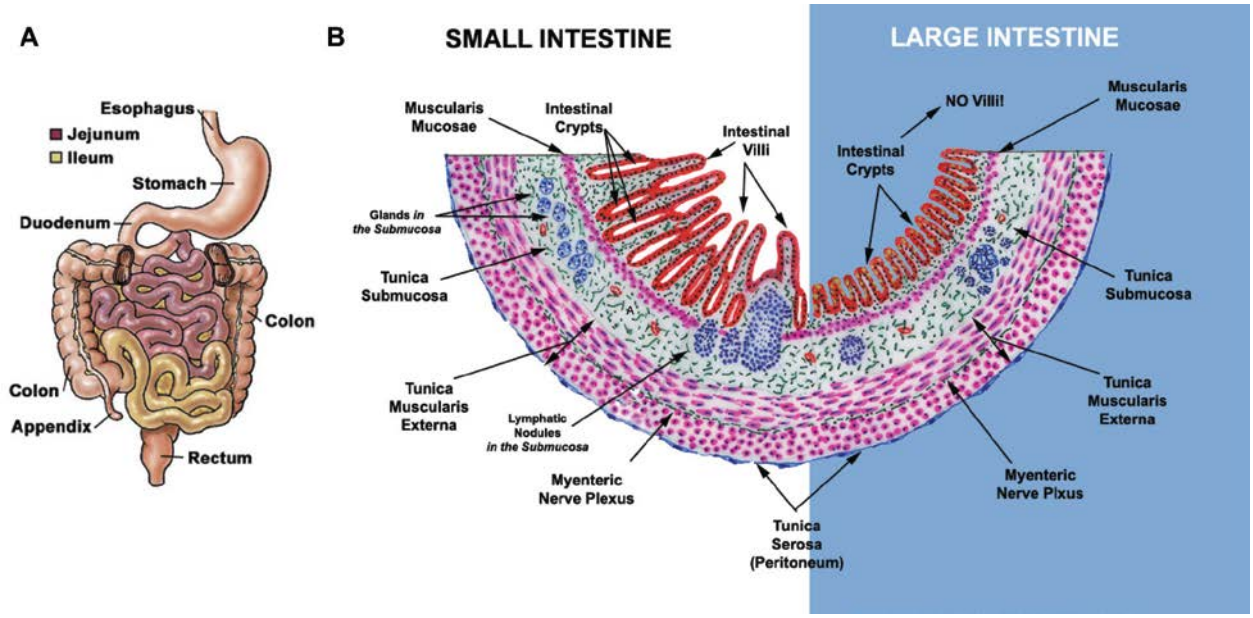



**Figure 1. The intestinal tract: Macro and micro structure**. A) Anatomy of the digestive tract (http://www.yoursurgery.com/ProcedureDetails.cfm?BR=1&Proc=49). B) Diagrammatic representation of a transverse section of the small intestine and the large intestine (http://www.vetmed.vt.edu/education/curriculum/vm8054/Labs/Lab19/Lab19.htm).

## 1.2 Major cell types of the intestine

Four different cell types mediate the major functions of the intestine, although at least seven cell types have been reported to exist (Gerbe et al., 2012), namely, enterocytes (absorptive cells), goblet (mucosecreting cells), enteroendocrine and Paneth cells (Figure 2). Enterocytes are by far the most abundant cell type in the intestinal epithelium as they constitute roughly 80% of the epithelium and are responsible for the absorption of nutrients and secretion of digestive enzymes. Goblet cells secrete mucus to protect the epithelium from physical/chemical assaults of luminal contents and their numbers vary considerably along the cephalocaudal axis, being increasingly abundant in the ileum and the colon. Enteroendocrine cells, representing less than 1% of epithelial cells, secrete hormones that regulate gastric and pancreatic secretion as well as intestinal movement. Several enteroendocrine subtypes have been described (Schonhoff et al., 2004). Paneth cells represent the only differentiated cell type that resides exclusively in the base of the crypt of the small intestine, intermingled with stem cells. They secrete antimicrobial agents and growth factors to neighboring stem cells and have been shown to constitute an important part of the intestinal stem cell niche (Durand et al., 2012; Sato et al., 2011). It was recently demonstrated that the colonic epithelium, devoid of Paneth cells, possesses a mucous secreting subpopulation of cKit positive Paneth-like cells (or deep crypt secretory cells) that apparently constitute part of

the colonic stem cell niche (Rothenberg et al., 2012). Often overlooked, two other cell types exist in the intestinal epithelium namely Microfold (M cells) and Tuft cells. M cells cover lymphoid tissue (Peyer's patches) and are believed to act as an antigen sampling system by facilitating the transport of antigens or microorganisms to underlying antigen-presenting cells (Neutra, 1998). Tuft cells are thought to play a role in inflammation by apparent secretion of prostanoids (Clevers and Batlle, 2013; Gerbe et al., 2012).

Terminally differentiated cell types are short lived as they enter an upward migratory flow to the surface epithelium (in colon) or tips of villi (small intestine), where they are thought to undergo apoptosis before being shed into the intestinal lumen (Figure 2). In the murine small intestine this process is reported to take 3-4 days while in the colon it takes 5-8 days (Karam, 1999). Paneth cells of the small intestine and the so called deep crypt secretory cells are reported to persist for 6-8 weeks and 14-21 days, respectively (Ireland et al., 2005; Karam, 1999). The lifelong continuous renewal of the intestinal epithelium is maintained by 14-18 intestinal stem cells (ISCs) located at the base of each crypt (Cheng and Leblond, 1974; Snippert et al., 2010). These ISCs undergo a single division every day while their rapidly dividing progeny, transit amplifying cells (TA cells), divide nearly twice as rapidly. TA cells reside in crypts for a period of 2-3 days and undergo up to 6 divisions before their upward migration progressively obligates them to commit to one of the six specialized lineages of the intestinal epithelium (Marshman et al., 2002).

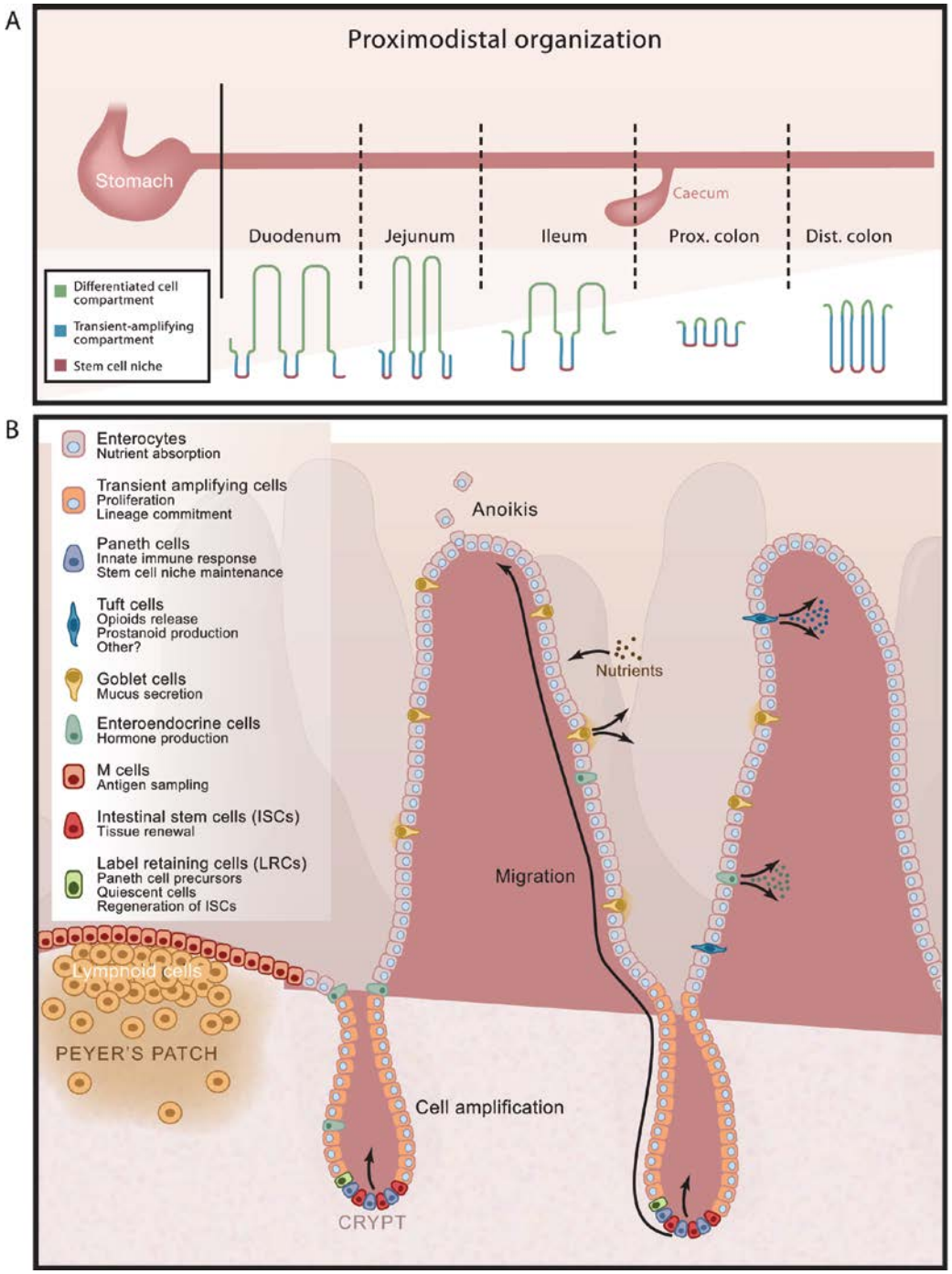



**Figure 2. The intestinal epithelium** (modified from (Clevers and Batlle, 2013). A) The intestine is anatomically divided into two main parts, the small intestine (duodenum, jejunum and ileum), which possess villi that progressively decrease in length and size, and the large intestine or colon (proximal and distal colon), which does not possess villi. B) ISCs are responsible for the lifelong renewal of the intestinal epithelium. ISC progeny, TA cells, enter the upward migratory flow to replenish the various terminally differentiated cell lineages of the intestine.

## 1.3 Intestinal stem cells: LGR5(+) stem cell zone model vs. +4 stem cell model

Adult stem cells are found in small numbers throughout the human body where they play a pivotal role in tissue maintenance. By definition, adult stem cells are multipotent and have the ability to self-renew. Multipotency refers to their ability to generate all cellular lineages of the given tissue, whereas self-renewal allows stem cells to maintain their own population. For over 30 years there has been a great deal of debate over the location of adult stem cells in the intestine. In summary, two prevailing models had been put forward, namely, the stem cell zone model (Cheng and Leblond, 1974) and the +4 model (Marshman et al., 2002; Potten, 1977).

### 1.3.1 Putative +4 ISCs

Back in the 70's, experiments of retention of labelled nucleotides such as 3H-thymthidine or BrdU lead by Christopher Potten identified a sparse population of cells located around crypt position +4 (i.e. immediately above Paneth Cells). These cells displayed two key features believed to be crucial for stem cells, namely radiation sensitivity to prevent expansion of damaged genetic information and DNA label retention as a result of either asymmetric segregation of new and old DNA strands (Potten, 1977) or due to a slow cycling/quiescence state . It was long accepted that these +4 label retaining crypt cells (LRCs) were ISCs. Over the last few years, multiple attempts to identify a quiescent/+4 label retaining stem cell population have led to the identification of several putative markers of this cell population namely, sFRP5, Dclk1 (also referred to as Dcamkl1), Wip1, Bmi1, Hopx, Tert and Lrig1 (Demidov et al., 2007; Giannakis et al., 2006; Gregorieff et al., 2005; Montgomery et al., 2011; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011; Van der Flier et al., 2007). Yet, evidence that the rare cells that express some of these markers around the +4 position are true ISCs is circumstantial as functional analysis such as lineage tracing are largely lacking.

### 1.3.2 LGR5(+) ISCs

The Clevers group (Barker et al., 2007) demonstrated that Lgr5 (Leucine-rich repeat-containing G protein-coupled receptor 5) is expressed in a small number of cells located at the crypt base intermingled with Paneth cells. Original work by Cheng and Berkens had previously named this particular cell type as crypt base columnar cells (CBCs). They also postulated that CBCs were ISCs.

Pioneering work by Clevers and colleagues demonstrated that LGR5(+) cells self-renewed and gave rise to all the lineages present in the intestinal epithelium by use of the $Lgr5^{eGFP}Cre^{ERT2}$ knock-in mouse model in combination with the $Rosa26R^{lacZ}$ mouse model used for lineage tracing (see section 2 [Figure 4]). In addition, they later demonstrated that these LGR5(+) cells could be indefinitely expanded *in vitro* in the form of intestinal organoids (Sato et al., 2009) and transplanted into superficially damaged mouse colon where they reconstituted a functional epithelial monolayer in vivo (Yui et al., 2012). Lgr5, which encodes a seven-transmembrane receptor is the most restricted ISC gene discovered to date (Munoz et al., 2012). LGR5 enhances WNT signaling by interacting with the WNT agonist RSPO1 (de Lau et al., 2011). All together, these studies provide strong support for the notion that CBCs (LGR5+) are bona-fide ISCs.

### 1.3.1.1 Other markers of the crypt base columnar/ISCs

Gene expression and proteomic analysis has led to the identification of other ISC restricted markers, although direct lineage tracing has only been performed with SMOC2 (Munoz et al., 2012) and indirect lineage tracing using the Lgr5 knock-in mice for the transcription factor ASCL2 (van der Flier et al., 2009b). Ascl2 is the only ISC specific gene to date that has shown to be essential for ISC maintenance, as its conditional deletion leads to rapid loss of ISCs, whereas overexpression results in ISC expansion (van der Flier et al., 2009b). SMOC2 has been described as a bone morphogenic protein (BMP) antagonist in the African clawed frog (Xenopus laevis) (Thomas et al., 2009) and since BMP inhibition is essential to maintain ISC cultures *in vitro* (Sato et al., 2009) it was thought to play a functional role in the ISC zone of crypts, although no intestinal phenotype was observed in mice lacking a functional Smoc2 allele (Munoz et al., 2012). Using a similar lineage tracing approach, another potential ISC marker, CD133, was shown to be expressed in ISC cells at the base of crypts (Zhu et al., 2009, Nature). Conversely, a similar knock-in model, independently produced by a second group, refuted the ISC specificity of CD133 (Snippert et al., 2009). Snippert and colleagues showed that CD133 preferentially marked cells located outside of the stem cell zone as well as demonstrated that a mere 10% of recombined crypts at 24h retained the tracing allele over time. Moreover, triple fluorescence *in situ* hybridization (ISH) confirmed that CD133 expression extended well beyond the stem cell zone (Itzkovitz et al., 2012). Another widely used ISC marker is OLFM4 (van der Flier et al., 2009a), although no mouse models have been published to characterize its specificity. Importantly, quantitative ISH showed that a number of so called +4 quiescent intestinal stem cell markers (see next section 1.3.2) including

Bmi1, Tert, Lrig1, Hopx and Olfm4 were enriched in LGR5(+) cells, yet possessed a broader expression pattern than LGR5 (Itzkovitz et al., 2012; Munoz et al., 2012).

### 1.3.3 Plasticity connects LGR5(+) ISCs and LRCs

A very recent study has shed light on quiescent intestinal stems by genetically marking slow proliferating cells in the intestinal epithelium as DNA label retaining cells (LRCs) and identifying that LRCs are in fact predominantly localized around position +3 and constitute roughly 20% of the LGR5+ cells in crypts (Buczacki et al., 2013). These LRCs persist for up to 4 weeks and their loss coincides with the concomitant increase in label retaining Paneth cells. Isolation and analysis of these LRCs allowed these authors to determine that they were Paneth cell precursors since they express both quiescent stem cell and Paneth cell specific genes. Upon injury, the label retaining Paneth cell precursors could be reverted to cycling ISCs, thus demonstrating the connections between these two stem cell pools. The Clevers group has also recently shown that Dll1(+) secretory precursor cells can revert to regenerate the LGR5(+)ISC pool upon tissue damage (van Es et al., 2012b). Targeted deletion of Lgr5 expressing cells from the intestinal epithelium using an elegant diphtheria toxin inducible allele demonstrated that the epithelium could be maintained for a period of at least 10 days in the absence of LGR5(+) cells (Tian et al., 2011). Importantly, these same authors show that a population of cells located above the LGR5(+) zone are capable of replenishing the ISC pool, after injury or retraction of diphtheria toxin. Taken together, all evidence points to the existence of plasticity in the intestinal crypt, whereby the immediate progeny of ISCs dedifferentiate upon injury to an ISC state to replenish the stem cell niche (Barker et al., 2012; Munoz et al., 2012; Tian et al., 2011) (Figure 3). The described studies mentioned above unify theories that both Cheng and Leblond (1974) as well as Potten and colleagues (Marshman et al., 2002) had postulated.

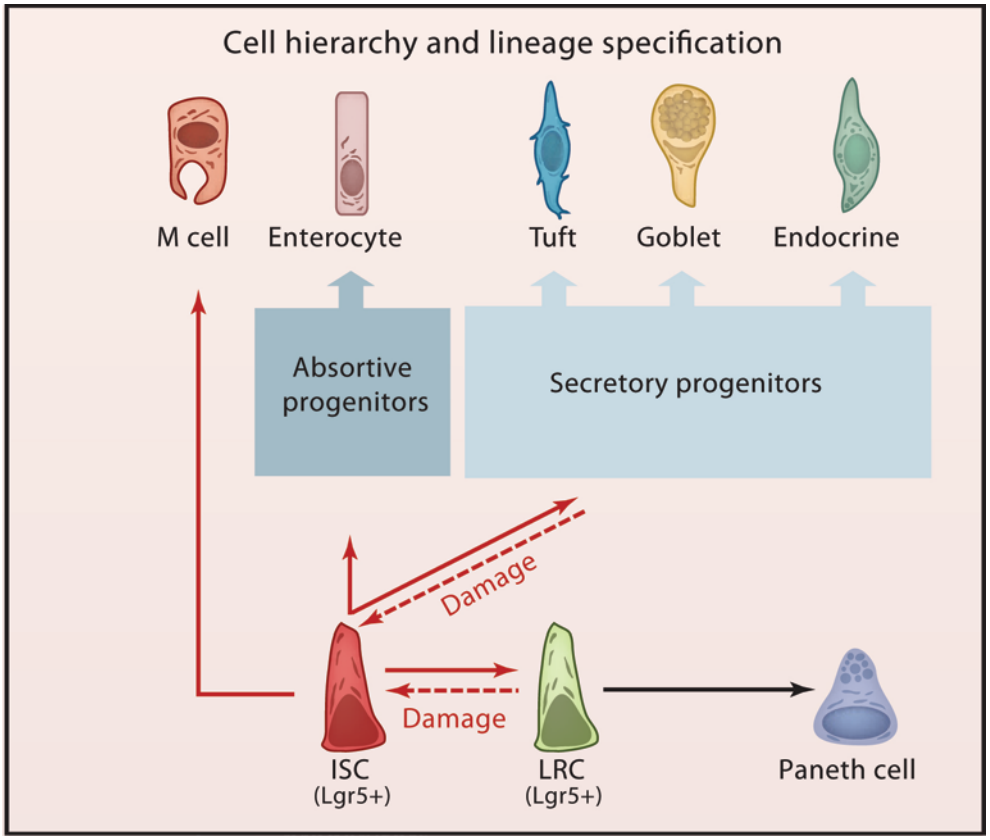



**Figure 3. Model of Cell Hierarchies and Plasticity** (adapted from (Clevers and Batlle, 2013). ISCs give rise to all cell types in the intestinal epithelium, including LRCs. Upon injury, which results in ISC loss, LRCs and early secretory progenitors can revert back to ISCs.

# 2. MOUSE GENETIC TOOLS USED TO STUDY INTESTINAL BIOLOGY AND DISEASE

## 2.1 Inducible Cre systems

Within the last 15 years great progress has been made with regard to genetic mouse models. Although a number of very interesting models exist, one that stands out and is currently widely used to study tissue specific biology is the CRE-ERT2 system. The original CRE-ERT2 construct was shown to be specifically activated by tamoxifen (4OHT) *in vitro* (Feil et al., 1997). Briefly, the ERT2 is the modified binding domain of the estrogen receptor, which is retained in the cytoplasm by heat shock proteins. Upon addition of tamoxifen the ERT2 undergoes a conformation change that allows it to translocate to the nucleus (Metzger et al., 1995). This construct was later used to make one of the most widely used murine intestinal specific Cre-loxP systems, the Villin $Cre^{ERT2}$ (el Marjou et al., 2004) to achieve very broad recombination of the great majority of

intestinal epithelial cells upon tamoxifen administration in vivo. This $Cre^{ERT2}$ has been used extensively to create other inducible tissue and cell specific mouse models for a variety of systems to activate or inactivate target genes of interest (Sauer, 1998). More recently, a bicistronic approach has been favoured whereby a fluorescent maker (e.g. GFP or other) is also knocked into the gene of interest in frame with the $Cre^{ERT2}$ to track cells of interest (Barker et al., 2007) (Figure 4). Another popular inducible intestine specific system is the lipophilic / xenobiotic inducible Cyp1a promoter (P450) Cre (Ireland et al., 2004), which is conditionally induced by treatment of mice with beta-Naphthoflavone (BNF) to activate the P450 promoter.

Cre expressing mouse lines are then crossed with mice possessing a gene of interest flanked by short genetic sequences termed loxP sites that are typically placed in non-coding regions (usually introns) surrounding exon(s) of the target gene of interest. The contents contained between the loxP sites are permanently excised in mice carrying the tissue or cell specific Cre, once animals are treated with either BNF or tamoxifen.

## 2.2 Lineage tracing

Lineage tracing is the identification of all progeny originating from a population of cells expressing a determined gene. In the intestine genetic lineage tracing in mice is typically performed using the Cre-loxP system, where Cre recombinase expression is driven by cell type specific promoters (e.g. putative stem cell markers). Cre expressing mouse lines are then crossed with mice possessing a reporter flanked by a loxP-STOP-loxP sequence. The STOP sequence flanked by loxP sites can be permanently excised in mice upon CRE activation (Figure 4).
When lineage tracing is desired, the reporter gene of interest is inserted into the ubiquitously expressed ROSA26 locus. The reporter construct typically used to trace cells include beta-galactosidase (Soriano, 1999) (LACZ), which has later been substituted for a myriad of florescent proteins (GFP, YFP, CFP, tdTomato, etc)(Figure 4).

## 2.3 Mouse models to study the initial steps of intestinal tumor formation

The prevailing model used to study tumorigenesis in mice is the Intestinal Neoplasia ($APC^{Min}$) mice (Oshima et al., 1996). These mice possess a truncating mutation in one copy of the tumor suppressor gene APC. Like FAP

patients (see section 3.1) these mice spontaneously lose their remaining wildtype APC allele (loss of heterozygosity), resulting in the development of numerous benign lesions (polyps) in the intestinal epithelium.

Another widely used model involves the conditional deletion of the tumor suppressor APC (Sansom et al., 2004) using the above mentioned Cre-loxP system. This system serves as a basis to study the immediate effects of APC loss and consequent aberrant WNT activation (see section 4.0) in a spacio-temporal fashion.

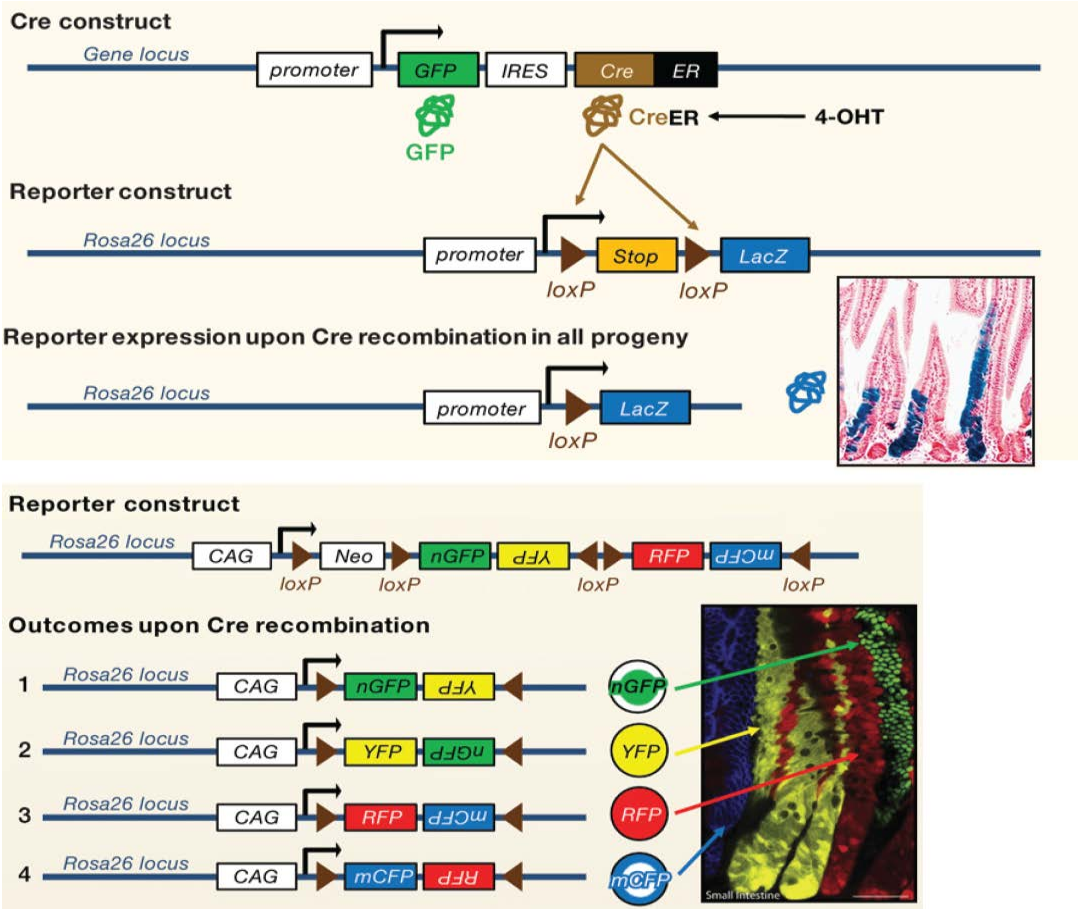



**Figure 4. Knock-in and lineage tracing in the intestine** (modified from (Kretzschmar and Watt, 2012).

# 3. COLORECTAL CANCER

## 3.1 Epidemiology and etiology

Colorectal cancer (CRC) is the second most diagnosed cancer worldwide in women and the third in men. It is estimated to be the third and fourth most

deadly cancer in men and women, respectively. A worldwide study for 2008 (GLOBOCAN) indicated that 1.2 million new cases have occurred with over 6 hundred thousand deaths and these numbers continue to climb (Jemal et al., 2011). Europe and North America are among the areas with the highest incidence rates. Moreover, CRC incidence rates are increasing at alarming rates in Spain and other nations with historically low rates of CRC (Jemal et al., 2011).

Approximately 75% of patients with colorectal cancer acquire the disease sporadically, with no apparent evidence of having inherited the disorder (Figure 5). Risk factors include older age, high fat, alcohol or red meat intake, obesity, smoking and lack of physical activity. Certain medical conditions also predispose individuals to develop CRC, particularly inflammatory bowel disease (ulcerative colitis or Crohn's disease). The sporadic appearance of adenomas is a common phenomenon as it has been estimated that by 70 years of age half of the population will develop one or more adenomas (Bond, 2000; Schatzkin et al., 1994), although only 10% of these lesions will progress to colorectal cancer (Levine and Ahnen, 2006).

Up to 25% of individuals with CRC have some family history that suggests a hereditary contribution (Burt, 2000; Kodach et al., 2008a; Lynch and de la Chapelle, 2003) (Figure 5). Compared to individuals with no family history, having a direct family member (parent, sibling or descendent) with CRC increases the risk of acquiring the disease by 2-3 fold. This risk doubles in cases in which more than one direct family member was diagnosed with the disease or they were diagnosed at a young age (Butterworth et al., 2006).

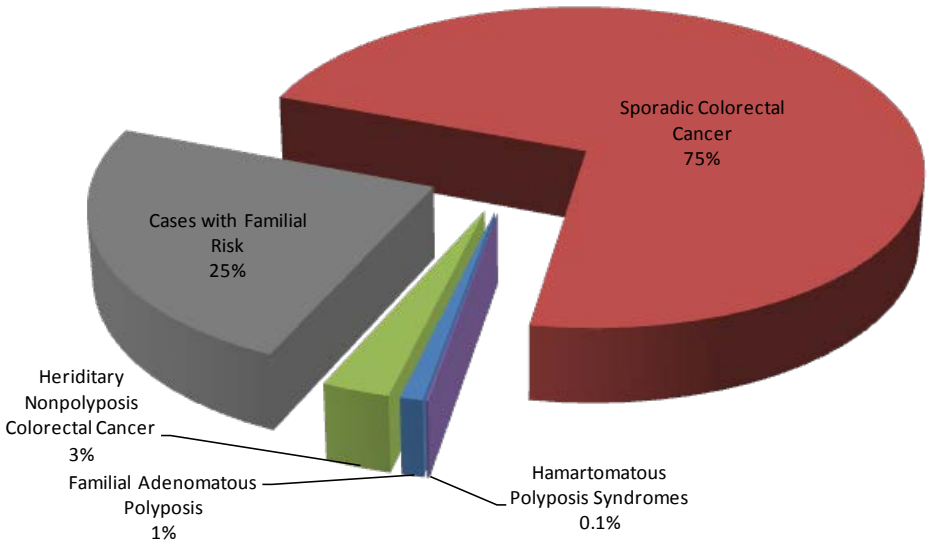



**Figure 5. Epidemiology of CRC** (modified from (Burt, 2000).

Only 5% of CRC are caused by known genetic mutations, namely, Hereditary Non-Polyposis Colorectal Cancer (HNPCC), Familial Adenomatous Polyposis (FAP), and Juvenile Polyposis Syndrome (JP) (Lynch and de la Chapelle, 2003). Almost all gene mutations known to predispose to CRC are inherited in an autosomal dominant form. Although inherited CRC syndromes are relatively rare, they have played a key role in defining the genetic pathogenesis of colonic polyps and cancers. HNPCC (also referred to as Lynch syndrome) is the most common of these syndromes (3%). Patients with HNPCC possess mutations in genes involved in mismatch repair such as MLH1 and MSH2. The inability to correct mismatches, deletions and insertions ultimately leads to microsatellite instability (MSI), which also occurs in 15-20% of sporadic CRC. Patients with HNPCC have an 80 risk of developing CRC (Vasen et al., 1996). FAP is the second most common inherited CRC syndrome (1%). FAP patients possess a truncating mutation in the Adenomatous Polyposis coli gene (APC), a key gene involved in the WNT signaling pathway (see section 4.0). The inherited mutant copy of APC induces the formation of adenomatous lesions once the normal inherited copy is lost. CRC risk in FAP patients approaches 100% (Jasperson et al., 2010). On average FAP patients are likely to develop CRC by 40 years of age, due to the burden of hundreds to thousands of colonic polyps. JP patients possess germline mutations in either BMPR1A or SMAD4, thus affecting BMP and/or TGFB signaling pathways (see section 6.0). (Hardwick et al., 2008; Howe et al., 2001). Unlike FAP and HNPCC, JP patients, as well as patients with Cowden Syndrome (CS) and Bannayan Riley Ruvalcaba Syndrome (BRRS), all develop mucin filled polyps surrounded by non-dysplastic tissue called hamartomatous polyps. Nonetheless, this rare autosomal dominant disease significantly increases the risk of developing CRC (up to 12-fold) (Brosens et al., 2007; Huang et al., 2000). Interestingly, the majority of juvenile polyps with wild type APC contain nuclear β-catenin (Iwamoto et al., 2005). Loss of function mutations in PTEN are found in patients with BRRS and CS (Eng and Peacocke, 1998; Waite and Eng, 2003). A link between BMP and PTEN has already been proposed (He et al., 2004). Although these inherited mutations can lead to CRC, it is speculated that their effects on the TGFB/BMP pathways have a more prominent role during later stages of tumor progression, namely the adenoma to carcinoma transition meaning the transition from benign to malignant (Hardwick et al., 2008) (see sections 3.2).

## 3.2 Pathogenesis

The vast majority of CRCs are initiated sporadically by the uncontrolled growth of epithelial cells that line the colon (Figure 6). This results in the formation of an excess number of epithelial cells that are disorganized and form glandular

clusters called adenomas. Once an adenoma has acquired sufficient mutational aberrations to become malignant it is referred to as adenocarcinoma, the most common type of colorectal cancer (95% of the cases). Other, rarer types include lymphoma and squamous cell carcinoma but will not be referred to in this work. The progression from benign lesions to fully malignant CRC is a process that takes many years, even decades (Kelloff et al., 2004; Kinzler and Vogelstein, 1996). Interestingly, various stages of the disease often co-exist at the time biopsies are taken and this has facilitated our understanding of the progression of the disease. The past few decades of research have identified specific genetic aberrations that occur in CRC. Genetic characterization of various stages of CRC allowed Fearon and Vogelstein to propose a model of successive genetic alterations in CRC, which continues to be used today (Fearon and Vogelstein, 1990). This model stipulates that colon tumors are formed due to genetic alterations that require several mutations to progress to malignancy and that, although a preferred sequence of these mutations exist, the accumulation of these mutations is the deciding factor in the tumor biology rather than the order in which the mutations appear.

The earliest detectable lesion leading to CRC is the formation of aberrant crypt foci (ACF) principally associated to aberrant activation of the WNT signaling pathway through truncating mutations in the tumor suppressor gene APC (>80% of the cases) (Figure 6) (Kinzler and Vogelstein, 1996). The mutations can be inherited or are acquired and likely occur in ISCs as shown in murine models (Barker et al., 2009). The pre-malignant lesions characteristically acquire further mutations through chromosomal instability (CIN). Over time, rapidly proliferating adenomas progress toward malignancy though the acquisition of further somatic mutations in other key molecular pathways, namely KRAS (mutations in *KRAS*), TGFB (mutations in *SMAD4* or *TGFBR2*) and P53 (mutation or allelic loss of *TP53*) (Fearon and Vogelstein, 1990) (Figure 6). More recently, somatic mutations that activate the PI3K/PTEN pathway have also been found in CRC (Markowitz and Bertagnolli, 2009).

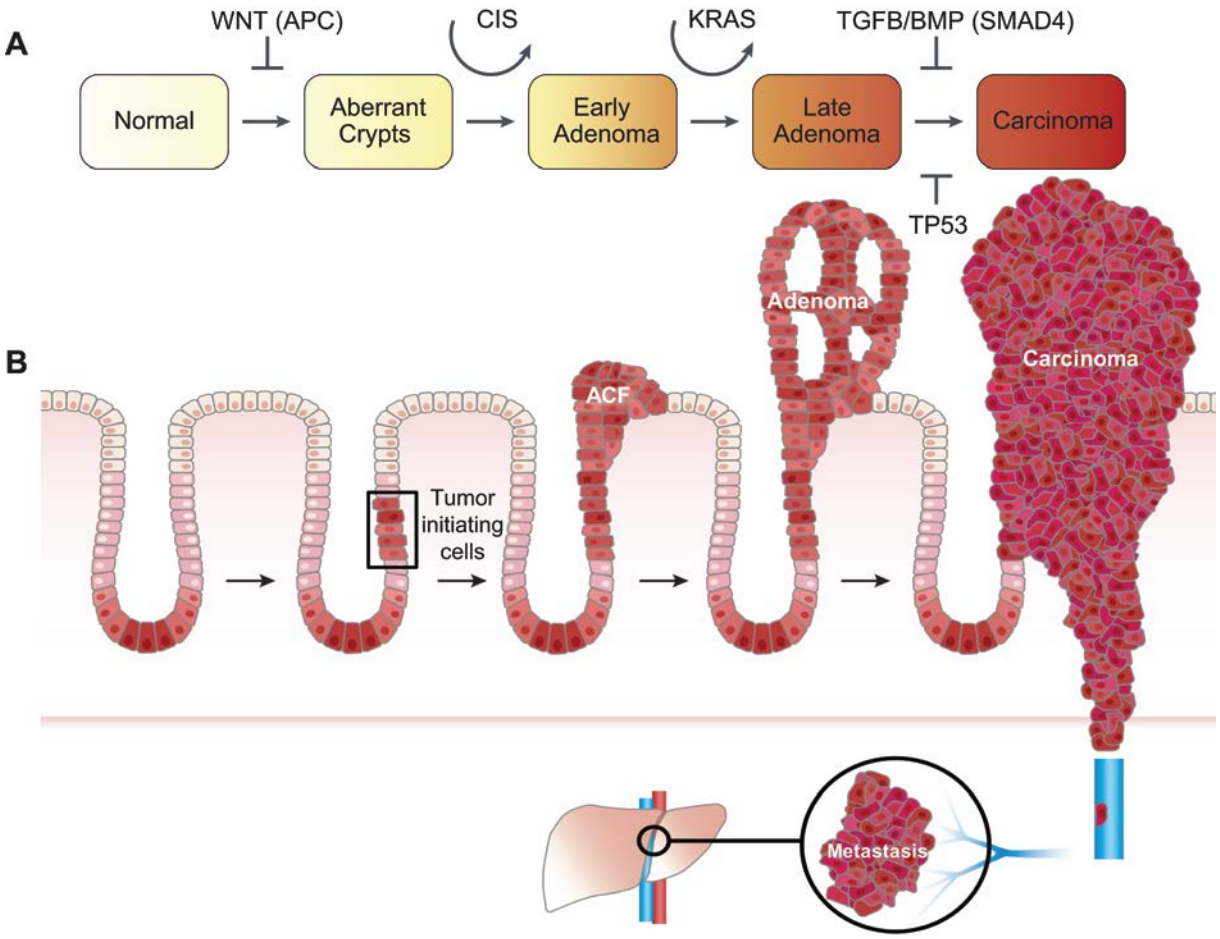



**Figure 6. Progression of CRC: Steps towards malignancy.** A. Modified version of the molecular events in adenoma to carcinoma sequence (adapted from (Fearon and Vogelstein, 1990; Hardwick et al., 2008). B. Scheme of events leading to malignancy (E. Batlle lab ©).

Importantly, a very recent study spearheaded by The Cancer Genome Atlas Network (TCGA) included whole-exome sequencing and integrative analysis of genomic data from 195 colorectal tumors confirming the key molecular pathways mentioned above (Cancer Genome Atlas, 2012). Figure 7 shows a summary of this study, including the frequency of mutations in components of the main signaling pathways (Figure 7).

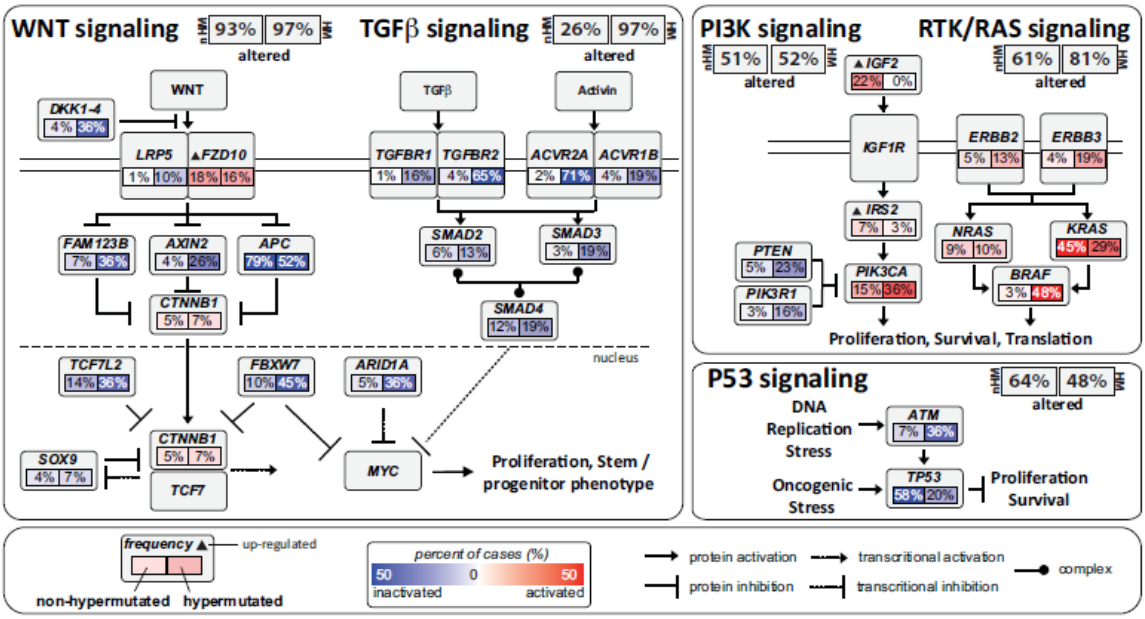



**Figure 7. TCGA Whole-exome sequencing and integrative analysis of CRC** (Cancer Genome Atlas, 2012). Patients were divided into two groups namely, microsatellite stable (e.g. non-hypermutated [nHM] n=165) and those with microsatellite instability (e.g. hypermutated [HM] n=30). Genes analyzed in the scheme possessed either mutations, deletions, amplifications of significant gene expression changes. Activated genes (red) and inactivated genes (blue).

## 3.3 Diagnosis, pathological classification and management

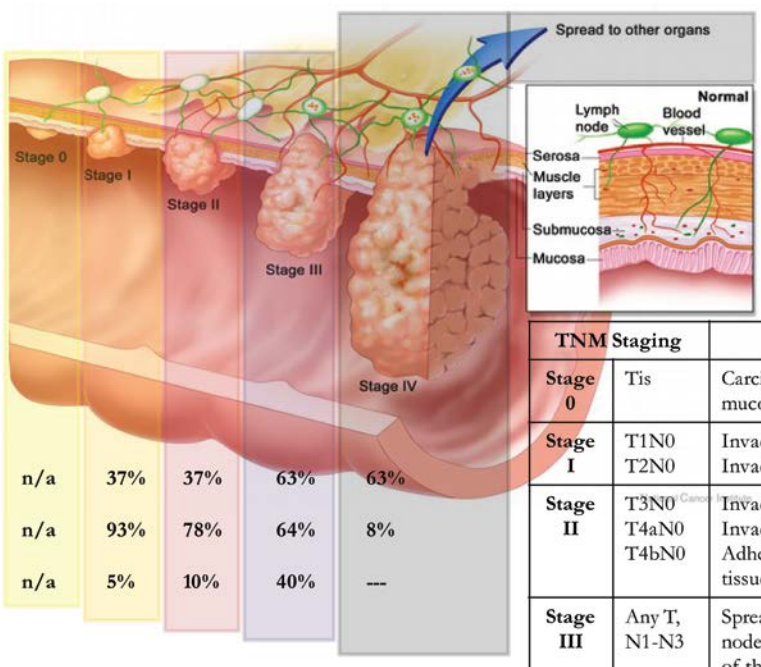
Diagnosis of colorectal cancer is typically performed by biopsy during colonoscopy or sigmoidoscopy. Once a patient is diagnosed with CRC, the degree of malignancy (or stage) of disease has to be determined. Using the combined information of a physical exam, biopsy and an array of advanced imaging techniques, clinicians can determine the clinical classification of the disease prior to surgical intervention. Surgery is typically performed once a diagnosis is made. At this point more detailed histopathological analyses can be performed to more accurately diagnose the stage of the disease. Pathological analyses of the tumor from a biopsy and/or surgery usually include description of tumor staging and grade. Tumor staging based on the TNM system describes the behavior and extent of spread of the malignant lesion.

### 3.3.1 TNM staging and disease management

The internationally accepted gold standard for staging of a given colorectal cancer lesion is set by the American Joint Committee on Cancer (AJCC). This staging system is referred to as the AJCC system or more commonly to the TNM (Tumor extent, Nodal involvement and presence of Metastases) system. The 7$^{th}$ edition of the TNM staging system is divided into 5 groups (Figure 8). Stage 0 is the earliest detectable malignant lesion, often referred to as "carcinoma *in situ*", which is confined to the mucosal layer. The treatment regimen can include removal using a colonoscope if confined to a single polypoid lesion (polypectomy), or the resection of a portion of the colon (colectomy) which often involves the surgeon to work in two parts of the abdomen at the same time to separate the colon from its blood vessels, resect the colon and rejoin the two remaining segments (anastomosis). Adjuvant therapy is not typically used. Stage I patients included tumors that have breached the mucosal layer, but have not yet reached the muscle layer. Treatment involves colectomy similarly to stage 0. Adjuvant therapy is not typically indicated. Stage II patients include tumors that have invaded the muscle layer (T3) or breached the serosa (T4a) and/or invaded adjacent tissues (T4b). Treatment involves colectomy. Adjuvant therapy includes chemotherapeutic agents alone or in combination with other drugs that enhance their efficacy and this treatment may be complemented with monoclonal antibody therapies. The benefits of using adjuvant therapy for stage II patients is currently quite controversial, as pointed out by the American Society of Clinical Oncology in collaboration with the Cancer Care Ontario Practice Guideline Initiative (Benson et al., 2004). Stage III patients include tumors that have spread to nearby lymph nodes (oval shaped structures that are ubiquitously distributed immune cell reservoirs linked to lymphatic vessels) but not to other parts of the body. Treatment involves local lymph node removal and colectomy using anastomosis. Adjuvant therapy includes chemotherapeutic agents alone or in combination with other drugs that enhance their efficacy and this treatment may be complemented with monoclonal antibody therapies. Stage IV patients include the dissemination of cancerous cells to distal organs, primarily to the liver and lungs. Depending on the gravity of the metastasis different treatment regiments may be pursued. Stage IV patients with resectable tumors are often given neoadjuvant chemotherapy to shrink the tumor prior to surgery. This may be followed by chemotherapy in combination with other drugs that enhance their efficacy and this treatment, as well as monoclonal antibodies and radiation therapy. Those stage IV patients with incurable metastatic lesions are treated as above, with the exception that surgical intervention, if performed, is palliative to increase quality of life.

## 3.3.2 Histological grading

While staging describes the extent to which the tumor has spread, histological grading is based on how similar the tumor appears to its tissue of origin (e.g. normal tissue). Grading systems are different for each type of cancer and different grading schemes exist such as: Four-tier (G1-G4), Three-tier (G1-G3) and Two-tier (G1-G2). The grading scheme will vary between institutions. Histological grading increases with lack of differentiation, typically going from well differentiated to poorly differentiated or undifferentiated cells. An example of a four tier grading is as follows: G1 (well differentiated), G2 (moderately differentiated), G3 (poorly differentiated) and G4 (undifferentiated). Tumors classified as poorly differentiated or undifferentiated have a worse prognosis than those deemed well or moderately differentiated. Similarly, samples classified as either high-grade (G3 or G4) have a worse prognosis than those with low grade (G1-G2). Grading can be used to decide whether adjuvant therapy should be applied.




| | | | | | |
|---|---|---|---|---|---|
| % diagnosed | n/a | 37% | 37% | 63% | 63% |
| 5 year survival | n/a | 93% | 78% | 64% | 8% |
| Disease relapse | n/a | 5% | 10% | 40% | --- |

| TNM Staging | | Tumor |
|---|---|---|
| Stage 0 | Tis | Carcinoma in situ, confined to mucosa |
| Stage I | T1N0<br>T2N0 | Invades through mucosa<br>Invades through submucosa |
| Stage II | T3N0<br>T4aN0<br>T4bN0 | Invades through muscle layer<br>Invades through serosa<br>Adheres to/invades adjacent tissue |
| Stage III | Any T, N1-N3 | Spreads to nearby lymph nodes but not to other parts of the body. |
| Stage IV | Any T, Any N, M1 | Cancer has spread to other parts of the body, such as the liver and lungs. |

**Figure 8. Overview of American Join Committee on Cancer TNM staging (7th ed.) for CRC.**

## 3.4 Prevention and prognosis

Preventive measures to detect polyps before they become malignant have a major impact on patient survival. Mortality rates can decrease significantly by performing colonoscopy and/or sigmoidoscopy screenings, typically after the age of 50. It has been estimated that colonoscopy has the potential to prevent approximately 65% of CRC cases (Brenner et al., 2007; Kahi et al., 2009). Moreover, sigmoidoscopy followed by colonosocopy, if a polyp or tumor is found, can identify the majority of the patients with advanced lesions (70-80%) and is associated with a 60-80% reduction in CRC mortality (Imperiale et al., 2000; Newcomb et al., 1992). Patient survival is intimately tied to TMN staging at the time of diagnosis as follows: stage I (93%), stage II (78%), stage III (64%) and stage IV (8%) (Kinzler and Vogelstein, 1996) (Figure 8). Not surprisingly, the probability of disease relapse parallels that of survival as follows: stage I (5%), stage II (10%), and stage III (40%). Disease relapse or cancer recurrence is classified as follows: local in which cancer cells are found in the original site and have not spread, regional in which cancer cells are found in the previous site and in nearby lymph nodes, but not in other parts of the body or distant in which cancer cells are found in a different part of the body presumable as a result of metastasis. Unfortunately, approximately 67% of patients are diagnosed at late stages (III-IV) of the disease.

# 4. WNT/BETA-CATENIN SIGNALING IN THE INTESTINE

## 4.1 WNT signaling pathway

The WNT signaling pathways are a group of signal transduction pathways that are activated by binding of WNT ligands to a Frizzled family receptor. The WNT signaling pathway that is referred to throughout this work is the so-called canonical WNT pathway that requires beta-catenin (βcat). The role of βcat independent non-canonical pathways, namely the planar cell polarity pathway and the WNT/calcium pathway, in normal intestinal homeostasis and disease is unclear and will not be discussed in this thesis.

Central to the WNT/βcat cascade is the stabilization/destruction of cytoplasmic βcat protein, which is tightly regulated by a complex of proteins collectively termed the “destruction complex”. In the absence of WNT ligand stimulation, βcat is recruited to the destruction complex by two scaffolding proteins: the tumor suppressor adenomatous polyposis coli (APC) and Axin. βcat is sequentially phosphorylated by the two other members of the complex: casein kinase 1-γ (CK1γ) and glycogen synthase kinase 3α/β (GSK3α/β) (Figure 9).

Phosphorylated βcat is then recognized and ubiquitinated (Ub) by the E3 ligase βTrCP within the destruction complex (Li et al., 2012). The proteosome then degrades ubiquitinated βcat in the complex maintaining low levels of βcat in the cytoplasm and nucleus (Li et al., 2012). Meanwhile, expression of WNT/βcat target genes in the nucleus is inhibited by association of transcription Factor 7-Like 2 (TCF4) or Lymphoid Enhancer Binding Factor 1 (LEF) with transcriptional repressors such as Groucho (Figure 9A).

Engagement of WNT ligands with the Frizzled (FZD) transmembrane receptor and the low-density-lipoprotein-related proteins 5 or 6 (LRP) co-receptor (Bhanot et al., 1996; Pinson et al., 2000; Tamai et al., 2000) results in activation of the WNT signaling cascade. Interaction of AXIN with phosphorylated LRP and Dishevelled (DVL) results in the inactivation of the destruction complex leading to accumulation of cytoplasmic βcat that translocates to the nucleus. In the nucleus, βcat binds TCF4 and/or LEF1 displacing the Groucho co-repressor and activates the transcription of WNT/βcat target genes (Waterman, 2004). Even though previously proposed models assumed either physical disassociation of the βcat destruction complex or impairment in βcat phosphorylation, it was recently shown that the destruction complex remains intact with phospho-βcat during WNT signaling (Li et al., 2012) (Figure 9B).

As mentioned above (section 2.2), the leading mechanism for CRC initiation is the mutation of *APC*. With a mutated *APC*, the destruction complex is no longer functional and β-cat accumulates to high levels in the cytoplasm, translocates into the nucleus and activates the transcription of WNT target genes (Figure 9C). While aberrant activation of WNT signaling in CRC is usually mediated by *APC* mutations, some cancers have increased β-cat because of mutations in β-cat itself that block its degradation, or they have mutations in other key genes in the signaling pathway namely as *AXIN1, AXIN2, TCF7L2*, or *NKD1*. The *APC* mutant colorectal cancer cells also maintain an intact destruction complex (Bao et al., 2012).

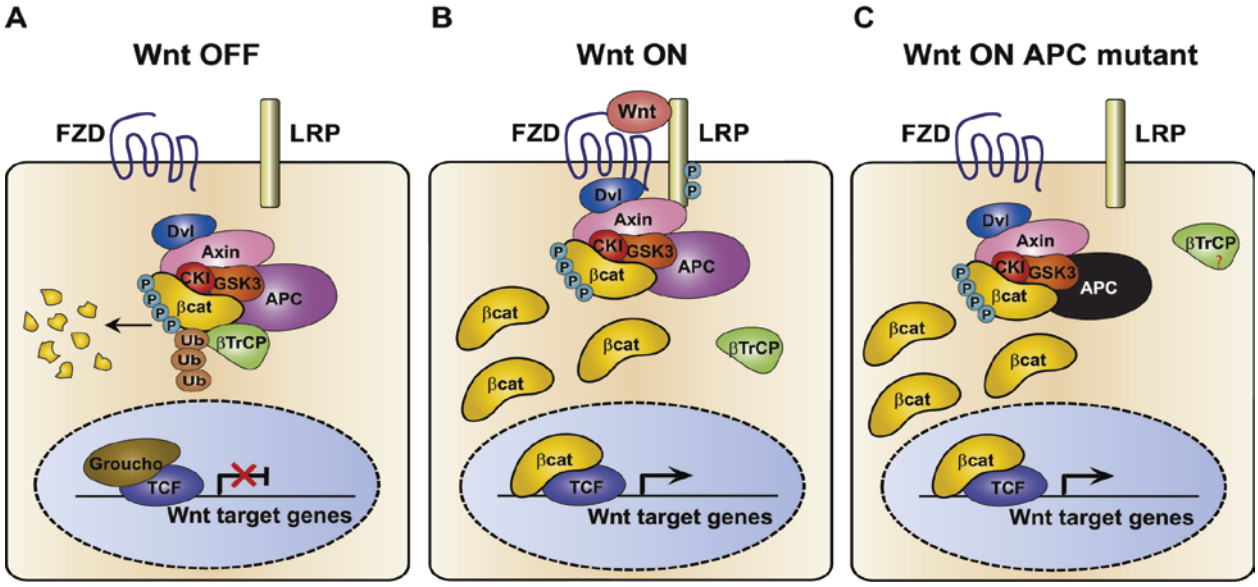



**Figure 9. WNT signaling pathway** (modified from (Li et al., 2012)

## 4.2 WNT beta-catenin signaling in homeostasis

The WNT signaling pathway is highly evolutionary conserved from fruit flies to humans (Nusse and Varmus, 1992). The WNT family of signaling proteins have been implicated in various processes both during embryonic development and adult tissue homeostasis (Moon et al., 2002; Wodarz and Nusse, 1998). In the mammalian intestine, canonical WNT signaling functions at the crypt base to regulate cell fate, proliferation and self-renewal of ISC and progenitor cells throughout life (Clevers, 2006; Korinek et al., 1998; Pinto et al., 2003; Radtke and Clevers, 2005; Sancho et al., 2004; van de Wetering et al., 2002a; van der Flier and Clevers, 2009). The normal intestinal crypt is subject to a gradient of extrinsic WNT signaling, that is strongest in the stem cell zone in the crypt base and lowest at crypt tops where terminally differentiated cells reside. Key findings performed by the Clevers group have shown loss of proliferating cells and subsequent attrition of the intestinal epithelium upon transgenic expression of the WNT inhibitor *Dkk1* (Pinto et al., 2003) or deletion of the WNT signaling transcription factor *Tcf7l2* (*Tcf4*) (Korinek et al., 1998). Recently, *Tcf1* and *Tcf3* were shown to be dispensable for epithelial homeostasis, confirming the requirement of *Tcf4* alone for intestinal homeostasis (van Es et al., 2012a). Conversely, activating mutations in βcat (Harada et al., 1999) or deletion of the tumor suppressor *Apc* (Sansom et al., 2004) resulted in hyperproliferation and adenoma formation. WNT ligands are at the apex of canonical WNT signaling and their expression patterns in the intestine have been shown to vary (Gregorieff et al., 2005). Epithelial cells in the bottom-most position of the crypts

have nuclear staining, suggesting expression must be spatially restricted. Recent work has shown that Paneth cells secrete WNT3 to the ISCs and thus constitute, at least in part, the WNT ligand source in the small intestine (Sato et al., 2011). However, the colonic WNT ligand source has not been elucidated. Since the discovery of the importance of WNT signaling in the intestine, a considerable interest has focused on identifying which downstream targets may be effectors of the pathway. To date, only a few of them have been investigated in the context of intestinal homeostasis or tumorigenesis.

## 4.3 WNT signaling in CRC

Deregulation of the WNT signaling pathway has been implicated in vast numbers of pathologies (Clevers, 2006; Moon et al., 2004). As mentioned above (section 3.2), the most commonly mutated pathway in CRC is the WNT signaling pathway (Cancer Genome Atlas, 2012) (Figure 7). The genetic program driven by βcat/TCF transcription in a CRC cell line was deciphered by genetically engineering CRC cells to inducibly block βcat/TCF activity (van de Wetering et al., 2002a). Upon analysis of a set of putative WNT target genes in mouse and human tissues, it was observed that both early CRC lesions and cells from the stem/progenitor compartment in the crypt base expressed a similar genetic program. This observation led to the proposal that the first step towards malignancy is the acquisition of crypt stem/progenitor cell phenotype (Batlle et al., 2002; van de Wetering et al., 2002a).

WNT target genes are highly upregulated in colorectal lesions compared to normal tissue (Van der Flier et al., 2007). The MYC oncogene is among the best examples in terms of critical downstream WNT targets in CRC. It has been shown that MYC plays a central role in the proliferative phenotype imposed by constitutive activation of the pathway (van de Wetering et al., 2002a). Blockade of WNT signaling in colorectal cancer cells results in a strong decrease in MYC levels followed by cell cycle arrest and differentiation toward enterocyte and secretory lineages. The original phenotype could be rescued by exogenous expression of *Myc*. The discovery of the location of ISCs and the increasing body of evidence suggesting that they constitute the cells of origin in CRC, has prompted the scientific community to further focus on those WNT targets restricted to this cell population, as they may constitute an ideal target to future therapies. It has been shown that WNT target gene expression occurs in 3 separate areas in the crypt namely, the proliferative compartment, ISC compartment and in terminally differentiated Paneth cells (Clevers and Batlle, 2006; Van der Flier et al., 2007). As mentioned above (section 1.3.1), the only confirmed ISC-restricted markers discovered to date are WNT target genes:

*Lgr5* (Barker et al., 2007), *Ascl2*, (van der Flier et al., 2009b) and *Smoc2* (Munoz et al., 2012). The identification of ISC specific markers has now made it possible to assess the role these cells might have once they have undergone oncogenic transformation.

### 4.3.1 Identification of the adenoma stem cell hierarchy

After the identification of LGR5(+) ISCs (see section 1.3.1), Clevers and colleagues pursued the issue of the cell of origin of intestinal cancer (Barker et al., 2009). To address this question they used genetic mouse models to deleted *Apc* either in ISCs (LGR5+) or in their progeny (TA or differentiated cells). Whereas deletion of *Apc* in LGR5(+) cells resulted in the formation of numerous adenomas, *Apc* deficiency in TAs rarely gave rise to large adenomas but instead generated small lesions confined to aberrant crypt foci (ACF). This led the authors to propose that LGR5(+) ISCs represent the cells of origin of intestinal cancer (Barker et al., 2009). Notably, despite high nuclear βcat (a hallmark of active WNT signaling) and homogenous MYC expression throughout the tumor mass, the vast majority of the transformed cells in the adenomas arising from LGR5(+) cells were negative for the expression of LGR5. Only few cells sitting at the base of tumor glands retained expression of LGR5(+). This observation raised the possibility that adenomas retained a hierarchical organization. To explore this possibility, authors performed lineage tracing experiments of LGR5(+)cells in an *Apc* null background using an elegant tracing allele, the multicolor Rosa26R-confetti (Figure 4) (Schepers et al., 2012). This tracing allele enables the fluorescent protein of the original trace to undergo a color change upon a second activation of the CRE recombinase by tamoxifen administration. These labelled LGR5(+) cells reconstituted the adenomas and all cell types within, while self-renewing their population. Remarkably, adenoma LGR5(+)cells displayed expression profiles similar to that of the normal ISCs (Barker et al., 2007).

Similar conclusions were obtained from a recent study showing that DCLK1 (Doublecortin-like kinase 1) marks adenoma stem cells, but not normal ISCs (Nakanishi et al., 2013). Nakanishi and colleagues generated mice in which a $Cre^{ERT2}$ was knocked-in the *Dclk1* locus. Lineage tracing of DCLK1(+) cells in the normal intestinal epithelium showed that this population were post mitotic tuft cells as previously reported (Gerbe et al., 2009). However, in adenomas DCLK1 marked a small proportion of LGR5(+) cells that had self-renewal capacity and that generated all the adenoma cell types. Remarkably, targeted deletion of *Dclk1* expressing cells from adenomas using an inducible diphtheria

toxin allele resulted in the collapse of established adenomas with little or no effect on surrounding normal tissue (Nakanishi et al., 2013).

# 5. COLON CANCER STEM / TUMOR INITIATING CELLS

The vast majority of tumors consist of heterogeneous cell populations identified by marker expression and growth capacity (Bonnet and Dick, 1997; Dalerba et al., 2007a). This heterogeneity is reminiscent of cellular hierarchies observed in normal adult tissues. This observation led to formulation of the Cancer Stem Cell (CSC) hypothesis(Dalerba et al., 2007a; Reya et al., 2001), which assumes that a small population of cells, the so-called CSC/Tumor initiating Cells (TIC), within the tumor bulk is clonogenic, possesses tumor initiating capabilities and is the origin from which all other more differentiated tumor cells arise (Dalerba et al., 2007a; Vermeulen et al., 2008) (Figure 10).

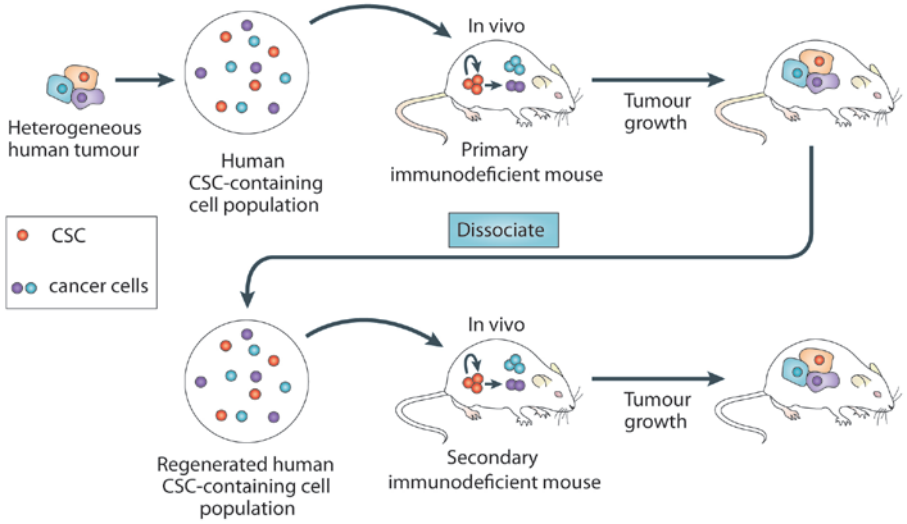



**Figure 10**. **Serial passaging of CSC/TIC from heterogeneous human tumor** (adapted from (Nguyen et al., 2012). CSC (orange) and more differentiated cancer cells (purple and blue). Only CSC are clonogenic, can initiate tumors and give rise to more differentiated progeny.

A significant number of studies have been performed to identify and isolate CRC-SCs using antibodies against differentially expressed surface markers, although the lack of CRC-SC specific antibodies has greatly hampered our ability to study the role of these cells in disease. This has led to the use of one or multiple markers to enrich for these elusive cells. The gold standard for validation of CSCs retrieval is performed by assessing the cells ability to initiate tumors that resemble the original heterogeneous tumor mass in immunodeficient xenograft models (Figure 10). In addition, the xenografts must

be capable of being serially passaged (Clarke et al., 2006; Dalerba et al., 2007a; Vermeulen et al., 2008). Nonetheless, the *in vitro* ability of prospective CSCs to form spheres that can be maintained over multiple passages has also been shown to be an effective surrogate to in vivo transplantation assays (O'Brien et al., 2012). Most studies have utilized differentially expressed cell surface markers to enrich for CRC-SCs, namely CD133 (O'Brien et al., 2007; Ricci-Vitiani et al., 2007), CD44 (Dalerba et al., 2007b), CD166 (Dalerba et al., 2007b) and EPHB2 (Merlos-Suarez et al., 2011). All of these markers were used to enrich a small fraction of these cells (often representing less than 1% of the total cells) by Fluorescence Activated Cell Sorting (FACS). Isolated cells were typically injected into immunocompromised mice where they were capable of initiating growth that, for the most part, recapitulated the heterogeneity found in the original tumor from which they were derived.

The pentaspan membrane protein CD133 (or Prominin-1) has been described as a marker of stem cells from the central nervous and hematopoietic systems (Uchida et al., 2000; Yin et al., 1997). More recently, it has been proposed as a CRC-SC marker (O'Brien et al., 2007; Ricci-Vitiani et al., 2007). Comparative analyses of CSC in CRC patient samples using CD133, CD44 and CD166 antibody stainings, alone or in combination, showed that CD133 was the best sole marker to predict patient survival (Horst et al., 2009b). Yet these markers did not always identify a rare population of cells, as some tumors possessed high levels of positivity (Horst et al., 2009b). Studies have also linked high expression levels of CD133 and nuclear βcat to vastly reduced survival rates (Horst et al., 2009a) and distant metastasis (Horst et al., 2009c; Neumann et al., 2012), although the role of CD133 in metastasis was unclear. However, using a mouse genetic model whereby LacZ was knocked into the *CD133* locus, it was shown that CD133 was broadly expressed in the intestinal epithelium (Shmelkov et al., 2008). Immunohistochemical analyses have further confirmed this finding in both mice and humans tissues, therefore, implying that CD133 is not restricted to CRC-SC (Karbanova et al., 2008; Shmelkov et al., 2008). Shmelkov and collaborators also showed that CD133(-) cell populations were able to form tumors in immunocompromised mice and that these tumors displayed a more aggressive phenotype than that of the CD133(+) engraftments (Shmelkov et al., 2008). Taken together, CD133 expression correlates with disease but its ability to identify CRC-SCs is questionable.

CD44 has been described as a CSC marker in breast (Al-Hajj et al., 2003), head and neck (Prince et al., 2007), prostate (Collins et al., 2005), pancreas (Li et al., 2007) and recently also the colon (Dalerba et al., 2007b). CD44 is a transmembrane protein that serves as a receptor for hyaluronic acid, which is involved in various cellular processes (reviewed in (Negi et al., 2012)). CD44

has been shown to be expressed in the proliferative compartment of normal intestinal crypts (Snippert et al., 2011). Dalerba and colleagues showed that CD44 alone enriched for CRC-SCs, yet the combination of both CD44 and CD166 (the activated leucocyte adhesion molecule [ALCAM]) enriched this population even further (Dalerba et al., 2007b). CD166 has been described as a mesenchymal stem cell marker in melanoma (van Kempen et al., 2000) and has been shown to be highly expressed in the ISC zone in mouse and human (Levin et al., 2010). However, CD166 expression has been reported to be a positive prognostic marker for overall CRC survival (Tachezy et al., 2012). Furthermore, CD44 and CD166 loss is associated with disease progression (Lugli et al., 2010).

The receptor tyrosine kinase EPHB2 is expressed in a gradient in the intestinal crypts with highest expression localized to the crypt base (Batlle et al., 2002). It plays a crucial role in cell positioning and migration as seen by loss of normal and tumor cell compartmentalization in knockout mice for this receptor or its ligand *EfnB1* (Batlle et al., 2002; Cortina et al., 2007). Our group recently used EPHB2 as surface marker to purify normal human colon stem cells for first time (Jung et al., 2011). $EPHB2^{high}$ human colon stem cells (CoSC) expressed a similar gene program to that of mouse ISCs and could be routinely grown in culture for extended periods of time (>2 months), while maintaining mutipotency. Our group also demonstrated that isolated $EPHB2^{high}$ cell populations from CRC patient samples expressed high levels of ISC genes compared to $EPHB2^{low}$ cells (Merlos-Suarez et al., 2011). Moreover, elevated expression of ISC genes in human CRC tumors positively associated to disease relapse. These $EPHB2^{high}$ tumor cells possessed high self-renewal and tumor initiating capacity when transplanted into mice. Thus, EPHB2 is a marker of both normal and CRC-SCs.

## 5.1 The paradoxical role of WNT signaling in CRC stem cells

It has been shown that high levels of WNT activity define a subpopulation of cells with the ability to self-renew and that possess tumor-initiating capacity (Kemper et al., 2012; Merlos-Suarez et al., 2011; Vermeulen et al., 2010). Moreover, these WNT high populations identify both colon stem cells (CoSC) and CRC-SCs. Gene expression signatures from the latter correlated with high risk of disease relapse (de Sousa et al., 2011; Merlos-Suarez et al., 2011).
In contrast to these findings, de Sousa and colleagues have also identified a subset of WNT target genes that, when silenced correlate with poor prognosis (de Sousa et al., 2011). This subset of WNT target genes included the ISC markers *LGR5* and *ASCL2* and other known WNT targets *APCDD1, DKK1* and

*AXIN2*. *In vitro* treatment of CRC lines and CRC-SCs with 5-aza-2'-deoxycytidine (AZA), a broad demethylating agent, allowed these authors to conclude that the underlying mechanism downregulating these WNT targets was promoter methylation. Previous work by our group had also shown that *EPHB2*, a WNT-target gene and bona-fide marker of CRC-SCs (Merlos-Suarez et al., 2011) is silenced as the disease progresses and tumors become more aggressive (Batlle et al., 2005). In fact, loss of *EPHB2* strongly correlated with degree of malignancy (Batlle et al., 2005). The proposed mechanisms involved in *EPHB2* silencing are promoter methylation and frame shift mutations (Alazzouzi et al., 2005), although others have claimed that *EPHB2* is not methylated but instead is regulated by the transcription factor REL (Fu et al., 2009).

# 6. TRANSFORMING GROWTH FACTOR BETA SUPERFAMILY

The Transforming Growth Factor Beta (TGFB) superfamily is composed of variety of secreted growth factors namely, TGFB and the bone morphogenetic proteins (BMP) as well as Activin (ACV), Inhibin (INH), Growth/Differentiation Factor (GDF), Anti-Mullerian hormone (MIS), NODAL and left-right determination factor (LEFTY), which are processed and secreted as homo or hetero dimmers (Chang et al., 2002). A wide array of developmental defects has been associated with mutation or loss of ligands, receptors or downstream mediators of this protein family (Aubin et al., 2004; Chang et al., 2002).
BMPs represent the largest subgroup of the TGFB superfamily. They are characterized by a high degree of promiscuity with regard to interactions with ligands and receptors of the TGFB superfamily (Kim and Choe, 2011; Mueller and Nickel, 2012). As many as twenty or so different BMP ligands have been reported, which accounts for nearly two thirds of the ligands in the TGFB superfamily (De Biase and Capanna, 2005; Derner and Anderson, 2005; Mueller and Nickel, 2012), unlike the simpler subgroups such as the TGFB subfamily that possesses only 3 ligands (TGFB1, 2 and 3) (Mueller and Nickel, 2012).

## 6.1 BMP signaling pathway

Dimeric BMP ligands (i.e. homo- or hetero- dimers) bind a heteromeric signaling complex composed of BMP receptor 1 (BMPR1) and BMP receptor 2 (BMPR2) (Figure 11). Three type 1 receptors have been shown to be bound by BMP namely, BMPR1A [ALK3], BMPR1B [ALK6] or Activin A receptor type 1 (ACVR1

[ALK2]) (Massague and Wotton, 2000). Similarly, three type 2 BMP receptors have been shown to be bound by BMP namely, BMPR2 or Activin receptor type2A or B (ACVR2A or ACVR2B) (Massague and Wotton, 2000). Both types of receptors are structurally very similar and are comprised of a short N-terminal extracellular ligand-binding domain, a single span-transmembrane domain and an intracellular domain with serine-threonine kinase activity. Importantly, an intracellular glycine/serine-rich domain (GS-box) is exclusive to type 1 receptors and is important for downstream signaling. Dimerization of the receptors results in the phosphorylation of GS box in the type 1 receptor, by way of the type 2 receptor. The activated type 1 receptor recruits and phosphorylates one of the SMADs, namely receptor SMAD1, SMAD5 and SMAD8. Phosphorylated receptor SMADs then form heteromeric complexes with the co-SMAD (SMAD4) and enter the nucleus where they interact with other transcription factors and regulate target gene expression such as the direct BMP targets inhibitor of DNA binding proteins (ID1, ID2 and ID3) (Langenfeld et al., 2013).

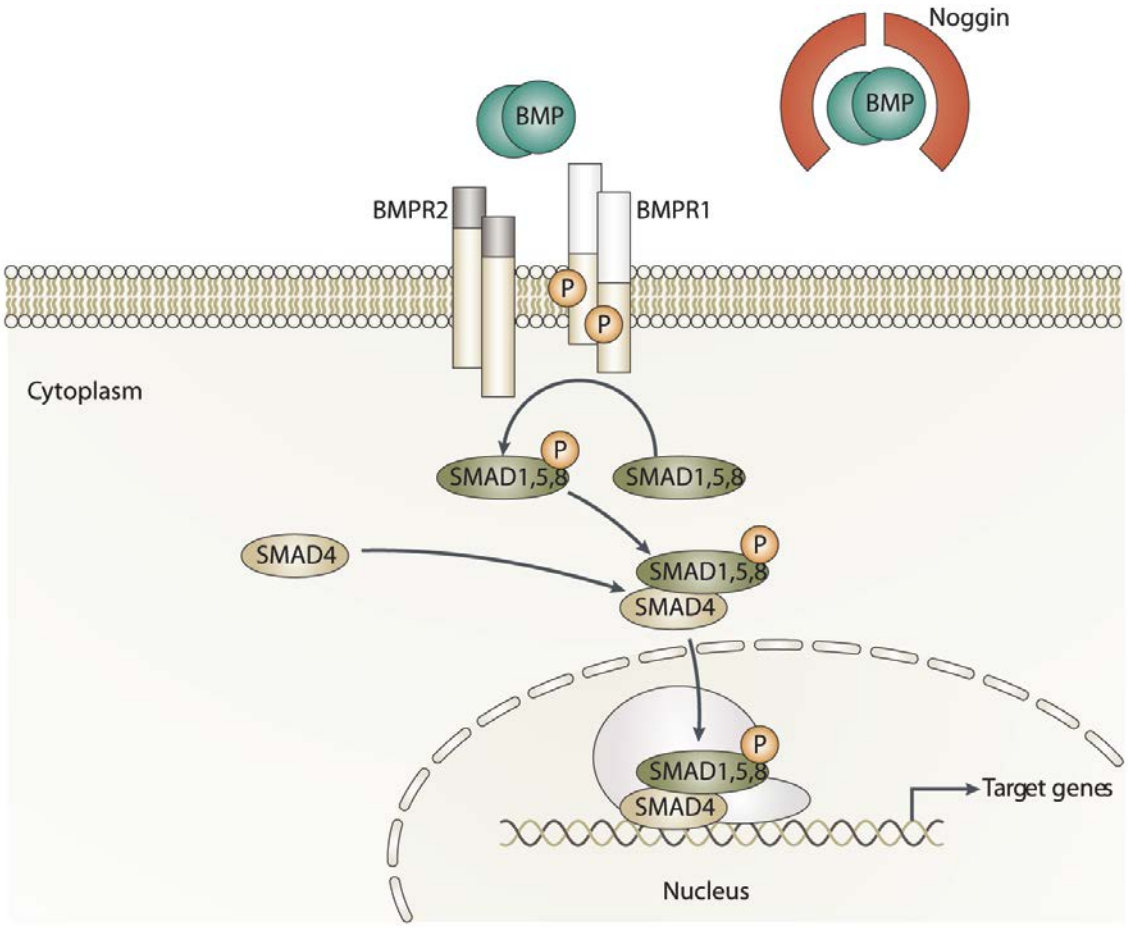



**Figure 11. BMP signaling Pathway** (adapted from (Hardwick et al., 2008))

## 6.2 Negative regulators of BMP signaling

BMP activity is regulated by several extracellular modulators including Noggin (NOG), Gremlins (GREM1 and GREM2), Chordin (CHRD), BMP and activin membrane-bound inhibitor (BAMBI) as well as intracellular modulators such as SMAD6, SMAD7 and SMAD ubiquitin Regulatory factors (SMURF1 and SMURF2).

NOG (Groppe et al., 2002; Groppe et al., 2003), GREM1 and 2 ((Pearce et al., 1999; Topol et al., 2000) and CHRD (Hyvonen, 2003; Larrain et al., 2000) directly bind and dimerize with BMP ligands effectively blocking signal transduction at the very apex of the pathway (Figure 11). BAMBI acts as a decoy receptor as it resembles the type 1 BMP receptors but lacks the GS box. As a result, it sequesters ligands from active receptors resulting in the inhibition of BMP signal transduction (Onichtchouk et al., 1999).

SMAD6 (Imamura et al., 1997) and SMAD7 (Souchelnytskyi et al., 1998) are inhibitory SMADs as they mediate degradation of type 1 receptors and SMADs (SMAD1, 2, 3, 5 and 8) from the BMP and TGFB subfamilies ((Cai et al., 2012; Souchelnytskyi et al., 1998; Zhang et al., 2007). SMAD7 blocks both BMP and TGFB signaling, while SMAD6 acts specifically on the BMP pathway (Zhang et al., 2007). SMURF1 and 2 are ubiquitin ligases that are recruited by SMAD6 and 7 to target type 1 receptors and receptor SMAD for proteasomal degradation (Murakami et al., 2003; Zhu et al., 1999).

The first synthetic BMP inhibitor, Dorsomorphin, was discovered by performing a compound screen using zebrafish (Yu et al., 2008b). Although it did not affect the TGFB pathway, it possessed many unwanted off target effects. This led to synthesis of two Dorsomorphin analogues with higher specificity and fewer off target effects namely, DMH1 (Hao et al., 2010) and LDN193189 (Yu et al., 2008a). Arguably the most specific BMP inhibitor, DMH1, has been shown to inhibit ACVR1 [ALK2] and BMPR1A [ALK3] activity but was reported to have negligible inhibitory effects on BMPR1B [ALK6] with no off target effects (Cross et al., 2011; Hao et al., 2010).

Apart from negative regulators of BMP signal transduction, there are also positive regulators often referred to as “co-receptors”, which modulate the interactions between BMP ligands and receptors. A number co-receptors have been reported to play a role in potentiating BMP signals largely through interactions with type 1 receptors namely, endoglin (END), TGFBR3 and RGMB (Kirkbride et al., 2008; Lee et al., 2009; Samad et al., 2005; Scherner et al., 2007).

## 6.3 BMP signaling in the intestinal epithelium

In the normal intestinal epithelium the study of BMPs has largely centered on BMP2, BMP4 and BMP7. *Bmp2* has been shown to be preferentially expressed in terminally differentiated epithelial cells on the intestinal surface of the colon or villi in the SI, whereas *Bmp7* was localized to myofibroblasts in the vicinity of the most terminally differentiated epithelial cells (Batts et al., 2006; Hardwick et al., 2004; Kosinski et al., 2007; van Dop et al., 2009). BMP4 has been shown to be expressed in mesenchymal cells located around crypts and within villi in the SI. (Batts et al., 2006; Haramis et al., 2004; He et al., 2004; Kosinski et al., 2007; van Dop et al., 2009). BMP signaling in the intestinal epithelium is typically assessed by immune detection of phosphorylated SMADs, namely activated SMAD1, SMAD5 and SMAD8 (Haramis et al., 2004; He et al., 2004).

BMP antagonists have been shown to be expressed in non-epithelial cells. In particular, NOG has been shown to be transiently expressed by stromal cells that surround the intestinal stem cells and in intestinal stem cells themselves (He et al., 2004). BMP modulators *GREM1* and *GREM2* have been shown to be expressed in myofibroblasts at the crypt base and in muscularis mucosae (Kosinski et al., 2007). Moreover the same authors showed evidence that treatment of normal intestinal cells with GREM1 enhances WNT signaling, as determined by cytoplasmic and nuclear translocation of βcat and increased WNT target gene expression. The expression patterns of BMP components fit with the idea that BMPs play a role in limiting proliferation and imposing differentiation.

After performing a comprehensive gene expression analysis of human crypt tops and crypt bottoms Kosinski and collaborators put forward a model whereby a decreasing BMP gradient from the crypt tops to the crypt bottoms opposes the WNT gradient emanating from the crypt bottoms to restrict stem cells expansion (Figure 12). They also propose that the WNT gradient is kept highest in the crypt base as a result of the underlying mesenchymal secretion of BMP antagonists that create opposing BMP signals to maintain WNT activity at the crypt base and promote stem cell self-renewal (Figure 12).

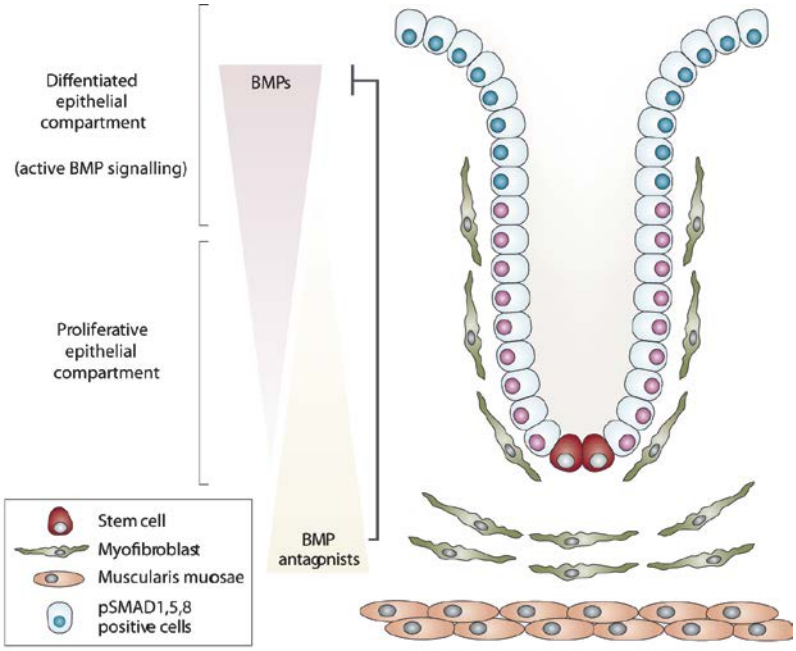



**Figure 12. Proposed model of the expression of components of the BMP pathway within normal colonic crypts** (adapted from (Hardwick et al., 2008; Kosinski et al., 2007)).

## 6.4 BMP in cancer (CRC)

As discussed in section 3.1, germline mutations in the BMP components SMAD4 and BMPR1A result in the formation of hamartomatous polyps that predispose patients to the acquisition of CRC. Several murine models have shown that BMP ligands have a tumor suppressive role in the intestinal epithelium. The transgenic overexpression of the BMP inhibitor NOG using the villin promoter was instrumental in demonstrating that abrogating BMP signaling in the murine intestinal epithelium, and consequently the stroma (as NOG is a secreted protein), resulted in the formation of hamartomatous polyps that phenocopied those observed in JP patients (section 3.1) (Haramis et al., 2004). Soon after, a second study performed by He and colleagues (2004) demonstrated that inducible expression of a mutant *BMPR1A* in both epithelial and mesenchymal cells lead to hamartomatous polyp formation (He et al., 2004). These two key studies demonstrated that mouse models could reproduce the onset of juvenile polyposis.

Other animal models have been used to examine the effects of BMP loss in different cell compartments. However, mesenchymal specific alterations in the BMP pathway have been shown to lead to hamartomatous polyps, whereas

epithelial specific deletions lack hamartomatous polyps, but instead lead to defects in epithelial differentiation. Whereas much debate exists regarding the specificity of the Cre drivers used (Auclair et al., 2007; Beppu et al., 2008; Kim et al., 2006), it seems clear that BMP pathway disruption in both epithelial and mesenchymal compartments is important for hamartomatous polyp formation. Mouse models harbouring alterations in BMP signaling suggest that BMP signals received by epithelial cells play an important role in differentiation and stem cell maintenance while BMP signals in the mesenchyme are crucial for crypt formation and restricting differentiation along the crypt axis. Future studies should undoubtedly shed light on the specific compartmental roles when BMP signaling.

Apart from its role in rare inherited diseases such as JP (<1% of CRC cases), the BMP pathway has been shown to be mutated, by loss of epithelial SMAD4 or BMPR2, in 70% of CRC (Kodach et al., 2008b). Histopathological analysis of both adenomas and carcinomas, also allowed Kodach and colleagues to conclude that loss of BMP signaling occurs at the adenoma to carcinoma transition. This suggests that loss of BMP signaling may have a more important role in tumor progression than tumor initiation, similarly to what was also been described for TGFB. This observation should not be entirely surprising as *SMAD4* was identified as a frequent mutation occurring in CRC, yet it has historically been associated to loss of tumor TGFB signaling and not loss of BMP signaling (Thiagalingam et al., 1996). As a result, the involvement of BMP signaling in late stage CRC is underrepresented in the literature when compared to the TGFB pathway.

# 7. GATA TRANSCRIPTION FACTORS

GATA are a family of evolutionarily conserved dual zinc finger transcription factors characterized by their ability to bind the core consensus sequence WGATAR (Maeda et al., 2005; Merika and Orkin, 1993). They are broadly divided into two groups based on expression patterns. GATA1-3 are largely expressed in hematopoietic cells (Orkin, 1998) while GATA4-6 are expressed in different endodermal and mesodermal tissues, namely the small intestine, colon, heart, stomach, liver, lungs, spleen, ovary, testis, and bladder (Kelley et al., 1993; Laverriere et al., 1994; Molkentin, 2000; Morrisey et al., 1996; Morrisey et al., 1997).

Knockout mice have been generated for *Gata4, Gata5* and *Gata6*. *Gata4* null mice die by embryonic day 9.5 (E9.5) and do not form the primitive heart tube and foregut (Kuo et al., 1997; Molkentin et al., 1997). *Gata5* null mice are born

fertile although they possess obvious abnormalities of the genitourinary tract of females (Molkentin et al., 2000). *Gata6* null mice die before gastrulation (Koutsourakis et al., 1999; Morrisey et al., 1998).

## 7.1 GATA factors in intestinal homeostasis

GATA4 and GATA6 are co-expressed in the proximal 85% of the small intestine. In the distal ileum, GATA6 is predominantly expressed in the crypts and mucous secreting cells while GATA4 is expressed in villi (Beuling et al., 2011; Bosse et al., 2006). GATA6 is the only GATA factor expressed in the colon. The constitutive knockouts precluded analysis of their functions in the intestinal epithelium in adult mice, as *Gata4* and *Gata6* null mice die before birth. Conditional *Gata4* (Bosse et al., 2006), *Gata5* (Laforest et al., 2011) and *Gata6* (Sodhi et al., 2006) knockouts using the Cre-loxP system have been generated to better dissect their roles during later stages of development and in adult tissues. Conditional deletion of *Gata4* from the intestinal epithelium using the intestinal specific driver Villin$^{Cre\text{-}ERT2}$ (el Marjou et al., 2004) resulted in a shift of intestinal identities along the cephalocaudal axis. More specifically, it has been shown to affect jejunal-ileal identities that include bile acid absorption in the proximal small intestine and decrease absorption in the ileum (Beuling et al., 2010; Bosse et al., 2006). Conditional deletion of *Gata5* has not been reported in the intestine, although heart tissue specific deletion leads to formation of bicuspid aortic valve, the leading congenital heart disease (Laforest et al., 2011). Conditional deletion of *Gata6* resulted in ileal defects in proliferation, Paneth cell differentiation, enteroendocrine cell commitment and enterocyte gene expression (Beuling et al., 2011). In the colon, *Gata6* loss resulted in defects in proliferation and mild defects in enteroendocrine sub-specification and colonocyte gene expression (Beuling et al., 2012).

## 7.2 GATA factors in cancer (CRC)

GATA6 is highly expressed in the vast majority of colorectal cancer cell lines, adenomas and adenocarcinomas (Haveri et al., 2008; Shureiqi et al., 2007). Few studies have analyzed its function in colorectal cancer. GATA6 has been proposed to promote colon cancer cell invasion by regulating urokinase plasminogen activator gene and prevent apoptosis by silencing 15-LOX-1 (Haveri et al., 2008; Shureiqi et al., 2007). Genomic mutations in GATA6 have been linked to congenital heart diseases (Lin et al., 2010; Maitra et al., 2010) and pancreatic agenesis (Lango Allen et al., 2012). In addition, GATA6 gene amplifications have recently been identified in pancreatobiliary and esophageal

cancers (Kwei et al., 2008; Lin et al., 2012). No *GATA6* genetic alterations have been found in CRC.

Interestingly, when it comes to GATA factor function, context seems to play a central role. Although GATA4 is considered a tumor suppressor in CRC, GATA4 levels are associated with poor prognosis in ovarian granulose cell tumors (Anttonen et al., 2005). Likewise, *GATA6* overexpression in CRC tumors and CRC cell lines suggests a role as an oncogene. Nonetheless, GATA6 is a bona-fide tumor suppressor in glioblastomas (Kamnasaran et al., 2007). Therefore, like with many other transcription factors, GATA6 function is entirely tissue and cell type dependent.

As stated above, GATA4 and GATA5 expression is absent in the colon. Interestingly, several studies have demonstrated that their silencing in CRC is due to promoter hypermethylation (Hellebrekers et al., 2009), although this observation seems questionable as they are not expressed in the colon. Accordingly, GATA4 and GATA5 have been shown to suppress colony formation, proliferation, migration, invasion and anchorage-independent growth of colorectal cancer cells, suggesting that they could play a tumor suppressive role (Hellebrekers et al., 2009).

# 8. HYPOXIA

Hypoxia is defined as a reduction in oxygen supply below physiological levels that restricts or even abolishes the function of organs, tissue or cells. The O2 levels that mark the hypoxic threshold vary considerably, although an upper limit of 35 mmHg (<5%) has been suggested (Hockel and Vaupel, 2001). Nonetheless, the use of 1% O2 is generally accepted and widely used in the field. Hypoxia is a common feature in a wide range of solid tumors, including colorectal cancer. Initially, it arises as a consequence of the rapid tumor cell proliferation and poor O2 diffusion as a result of disorganized intratumoral vascular networks (Jain, 2005; Pries et al., 2009). Hypoxia leads to a vast array of changes in tumor biology (Semenza, 2003) including pro-survival changes that suppress apoptosis (Erler et al., 2004), switch to anabolic metabolism (Cairns et al., 2011), loss of genomic stability (Guzy et al., 2005; Hill et al., 2009), epithelial to mesenchymal transition (EMT) (Hill et al., 2009), invasiveness (Pennacchietti et al., 2003) and metastasis (Chang et al., 2011). Indeed, a recent review by Ruan and colleagues argues that hypoxia alone mediates and regulates each of the hallmarks of cancer (Ruan et al., 2009). In addition to cellular responses, hypoxia typically associates with poor drug delivery and impaired response to traditional cancer therapies including

chemotherapy and radiotherapy due to restricted access to tumors (Vaupel and Mayer, 2007).

Anti-angiogenic therapy is typically used in combination with cytotoxic therapies for the treatment of cancer to prevent tumors from developing their own blood supply system that is required for growth. Antiangiogenic therapy has shown to delay tumor progression and improve overall survival compared with standard therapy in some patients, but results are modest (Rapisarda and Melillo, 2012). Nonetheless, such therapies increase intratumoral hypoxia in a variety of cancers, including colorectal cancer. Thus, it is no surprise that intratumoral hypoxia is associated with drug treatment failure and patient mortality (Semenza, 2007).

## 8.1 Hypoxia-inducible factor 1 (HIF1)

The main mechanism by which cancer cells adapt to hypoxia stress is through stabilization of the Hypoxia Inducible Factor 1 (HIF1) transcription factor, a heterodimer composed of an oxygen sensitive alpha subunit (HIF1A) and a constitutively expressed oxygen insensitive beta subunit (HIF1B) (Wang et al., 1995). HIF1A and HIF1B are ubiquitously expressed (Wenger et al., 1997). HIF1 is widely considered to be the master regulator of the adaptive response to hypoxia (Semenza, 1998). To date, HIF1 has been shown to transcriptionally regulate over >200 hypoxia responsive genes of which at least 70 have been shown to be directly regulated by HIF1 (Semenza, 2003, 2007; Takenaga, 2011). Figure 13 summarises a subset of these targets and their functions.

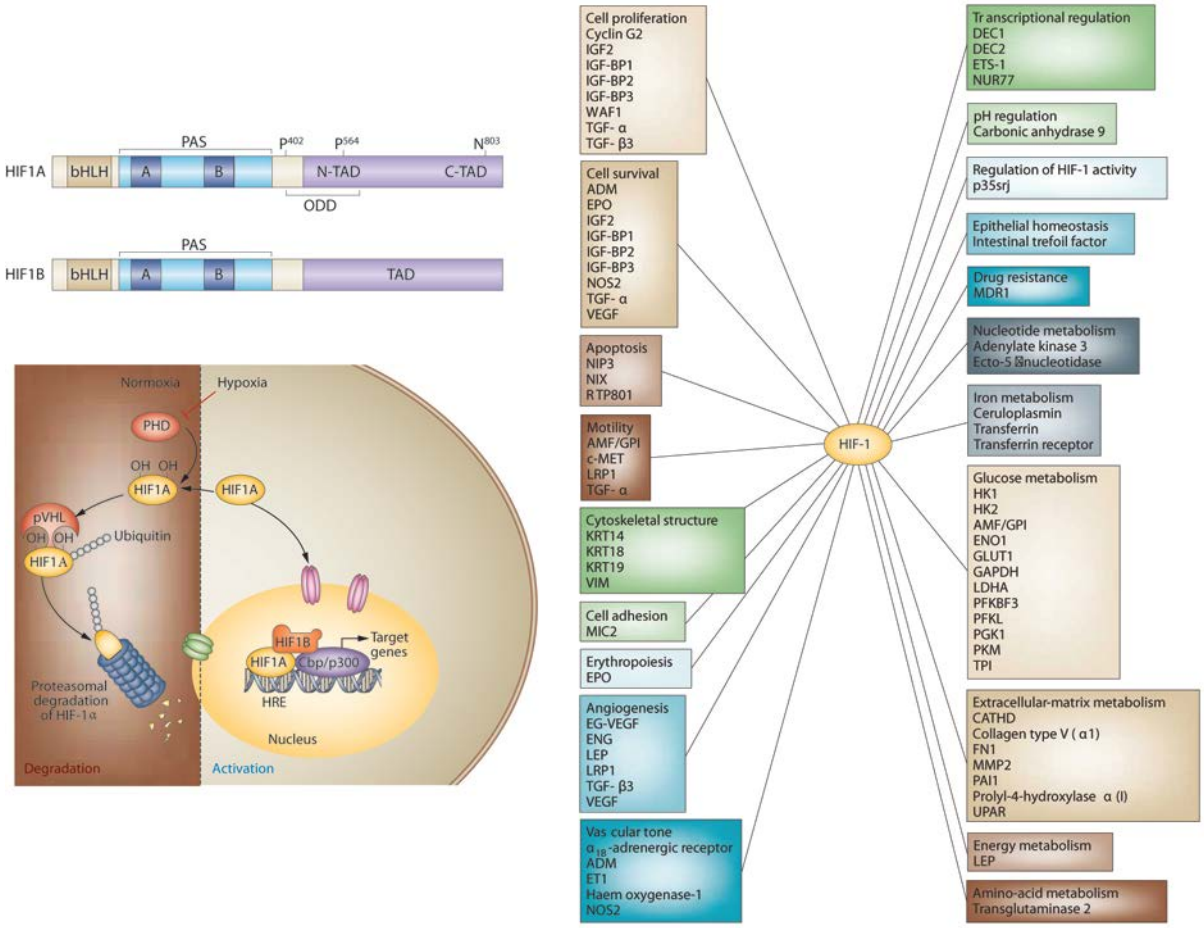



**Figure 13. Degradation and activation of HIF1A and downstream targets** (adapted from (Keith et al., 2012; Maes et al., 2012; Semenza, 2003)).

## 8.2 Regulation of HIF1A

In normoxia, HIF1A is subjected to prolyl hydroxylation at two proline residues (i.e. Pro402 and Pro564) located in the oxygen dependent degradation domain (ODD), as a result of its detection by prolyl hydroxylases (PHD1-3) (Figure 13) (Ivan et al., 2001; Jaakkola et al., 2001; Yu et al., 2001). The hydroxylated prolines allow recruitment of the von Hippel Lindau (VHL) tumor suppressor protein, a recognition component of the E3 ubiquitin ligase complex, resulting in polyubiquitination and proteosomal degradation of the HIF1A subunit (Maxwell et al., 1999; Salceda and Caro, 1997) (Figure 13). In tissues where the oxygen tension is >5%, the half-life of HIF1A is very short (<5 min) (Semenza, 2007). Additionally, another oxygen sensitive inhibitor of HIF1A, is the Factor Inhibiting HIF1 (FIH1), which hydroxylates Asn803 located in the c-terminus of HIF1A inhibiting its interaction with transcriptional co-activators p300 and CREB

binding protein (CBP) (Hewitson et al., 2002; Lando et al., 2002; Mahon et al., 2001).

Under hypoxic conditions PHD activity is inhibited (Kaelin and Ratcliffe, 2008) and HIF1A it is no longer targeted for proteosomal degradation and heterodimerizes with the HIF1B to activate target gene expression by binding with co-activators CBP/p300 at genomic regions containing HIF Responsive Elements (HRE) as summarized in Figure 13. Importantly, it has been reported that increased oncogenic signaling in cancer cells can induce the expression of the HIFA subunits through oxygen independent mechanisms (Pouyssegur et al., 2006).

# 9. EPITHELIAL TO MENSENCHYMAL TRANSITION (EMT)

Epithelial and mesenchymal cells have been historically identified on the basis of visual appearance and morphology. Epithelial cells typically form a highly organized monolayer of cells with tight cellular junctions that hold the epithelial monolayer together. These cell to cell contacts are largely mediated through adherent junctions, composed of cadherin and catenin proteins, which serve to link the actin cytoskeleton of neighboring cells together. These epithelial sheets typically possess apico-basolateral polarity as a result of their cell to cell contacts and association with the basement membrane, which gives them a certain degree of rigidity. Consequently, epithelial cells possess limited ability for migration. Conversely, mesenchymal cells do not associate with the basement membrane, lack apico-basolateral polarity and only make focal contacts with neighboring cells. They are typically spindle shaped and possess a front to back end polarity. As a result, they do not form organized monolayer structures. Their lack of cell to cell contacts affords them the ability to migrate as single cells, unlike their epithelial counterparts.

Epithelial to mesenchymal transition (EMT) is a biological process that endows a once polarized and rather immobile epithelial cell with a highly motile mesenchymal cell phenotype (Kalluri and Neilson, 2003). The completion of EMT requires a wide array of changes to occur in cells. These changes include the activation of key transcription factors, de novo expression of mesenchymal and repression of epithelial surface proteins, reorganization of the cytoskeleton and secretion of extracellular matrix (ECM) degrading proteins (Figure 14A). Many of the markers depicted in Figure 14 are bona-fide markers of EMT that have been widely used.

## 9.1 Overview of the three types of EMT

Recently, EMT has been divided into three subtypes based on distinct biological settings and functional consequences (Kalluri and Weinberg, 2009; Thiery et al., 2009). Type 1 EMT occurs as a result of cellular plasticity during early embryonic development and is required for the formation of new tissues and organs (Figure 14B). The primary epithelium originating from the endoderm undergoes EMT to give rise to the primary mesenchyme that migrates through the primitive streak to form the mesoderm or undergoes mesenchymal to epithelial transition (MET) to form the third germ layer, the mesoderm (Kalluri and Weinberg, 2009; Thiery et al., 2009).

Type 2 EMT occurs after development as a result of tissue regeneration after injury. Upon inflammation, signals derived from inflammatory cells and fibroblasts in the intestine, heart, liver, lungs, and kidneys activate a transcription program in epithelial cells that results in the expression of mesenchymal markers (Okada et al., 1997; Rastaldi et al., 2002). This process initially leads to a partial EMT, whereby epithelial cells express both mesenchymal and epithelial markers (Figure 14A). Ultimately, this will lead to a full EMT, but once inflammation has stopped cells typically revert back to an epithelial state (Arnoux et al., 2008; Choi and Diehl, 2009). When the liver is injured, EMT plays a role in promoting organ repair. Hepatocytes that have undergone EMT repopulate the organ and are reverted back to hepatocytes upon cessation of the inflammatory response (i.e. wound is healed), by way of MET (Choi and Diehl, 2009) (Figure 14B). However, this process can go awry and lead to fibrosis if the inflammation does not cease and becomes chronic. During fibrosis mesenchymal/fibroblastic cells accumulate in the organ and secrete large amounts of collagen, which accumulates and ultimately leads to organ failure. Thus, the physiological wound healing function of type 2 EMT can ultimately lead to a pathological condition (Thiery et al., 2009). Clear examples of EMT have been reported in patients with kidney fibrosis (Rastaldi et al., 2002) and Crohn's disease (Bataille et al., 2008).

Type 3 EMT occurs in epithelial cancer cells and plays a key role in cancer progression. EMT in the context of cancer implies that epithelial cells acquire the abilities to invade, escape apoptosis and eventually disseminate (Barrallo-Gimeno and Nieto, 2005; Polyak and Weinberg, 2009). This is in line with the observation that tumor cells at the invasive fronts of primary tumors lose expression of epithelial cell adhesion proteins, dissociate from neighboring epithelial cells and become single motile cells (Brabletz et al., 2001) (Figure 14B). The acquisition of the EMT phenotype by cancer cells is widely considered as a critical step in malignancy (Thiery, 2002) and the expression of

mesenchymal markers for *in vitro* and in vivo cancer models has been reported (Yang and Weinberg, 2008). In human tumors, EMT related mesenchymal markers are typically observed at the invasion fronts of many types of cancer, including colorectal cancer (Kahlert et al., 2011; Spaderna et al., 2006). It is widely accepted that the invasion front harbours the cancer cells that ultimately enter the metastatic cascade that end in the colonisation of distal organs (Brabletz et al., 2001; Fidler and Poste, 2008; Thiery, 2002), yet, metastatic lesions histologically resemble their corresponding primary lesions. Thus, it assumes that an MET must occur once the disseminated cells have reached their target distal organ (Figure 14C).

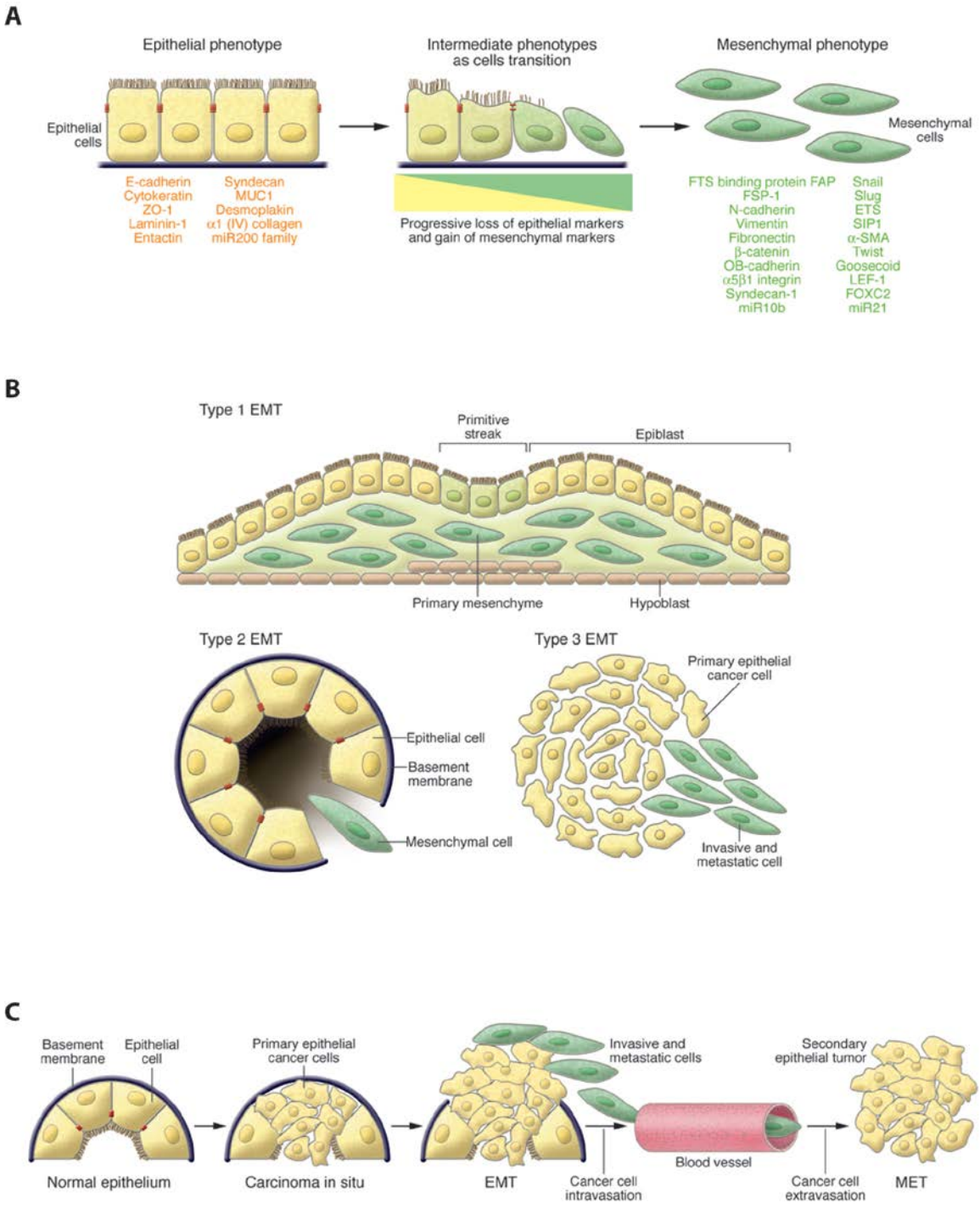



**Figure 14. EMT: the functional transition, types and involvement in cancer** (adapted from (Kalluri and Weinberg, 2009).

## 9.2 Pathways and factors involved in EMT

Many signaling pathways have been reported to be involved in regulating EMT. These include the WNT (Vincan and Barker, 2008), BMP (Bailey et al., 2007), TGFB (Mani et al., 2008), NOTCH (Bailey et al., 2007) and Sonic Hedgehog (SHH) (Bailey et al., 2007), as well as Receptor Tyrosine Kinases (RTK) pathways (Gentile et al., 2008). In addition, physiological mechanisms such as hypoxia (Gort et al., 2008) and stromal cell interactions (Sheehan et al., 2008) have been reported to induce EMT. These pathways ultimately converge on a limited set of transcription factors that drive EMT namely, TWIST1, TWIST2, GSC, TCF3(E47) (Slattery et al., 2006), TCF4 (E2.2) (Sobrado et al., 2009), KLF8 (Wang et al., 2007), FOXC2 (Mani et al., 2007), SNAI1, SNAI2 (Nieto, 2002), ZEB1 (Grooteclaes and Frisch, 2000), ZEB2 (Moreno-Bueno et al., 2008; Thiery et al., 2009). Many of these transcription factors have been shown to induce EMT during development and disease (Moreno-Bueno et al., 2008; Peinado et al., 2007). Also, microRNAs (miR141, miR200a, miR200b, miR200c, miR 429) have been shown to activate EMT (Gregory et al., 2008; Thiery et al., 2009). An important downstream consequence of the activation of these factors is the transcriptional control of the E-cadherin (*CDH1*) promoter. This is achieved by indirect or direct binding to E-box consensus sequences (Batlle et al., 2000; Cano et al., 2000). The latter would be the case of SNAIL and ZEB factors together with E47 and KLF8. There is a growing list of new inducers of EMT that will undoubtedly aid in understanding this complex process in both development and disease.

## 9.3 Involvement of EMT in CRC invasion fronts

In CRC, budding tumor cells (i.e. single or small clusters of 5 cells or less detached from the tumor bulk) located at invasion fronts may be histologically considered a mark of EMT (Prall, 2007; Zlobec and Lugli, 2010) and are reported to be present in 20-30% of tumors (Guarino et al., 2007; Ueno et al., 2002). Grade or de-differentiation (or tumor cell budding) is inversely correlated with clinical prognosis (Hostettler et al., 2010; Jass et al., 2003; Lugli et al., 2012). The process of tumor cell budding and migration through the stroma is widely considered an early and essential step leading to metastasis (Kevans et al., 2011; Zlobec and Lugli, 2010). Bud presence is predictive of lymph node metastasis, distant metastasis and local recurrence (Zlobec and Lugli, 2010). The molecular, genetic and physiological events triggering tumor budding are active areas of investigation.

# CHAPTER 2

# THE TRANSCRIPTION FACTOR GATA6 ALLOWS SELF-RENEWAL OF COLON ADENOMA STEM CELLS BY REPRESSING BMP GENE EXPRESSION.

Gavin Whissell[1], Paola Martinelli[2*], Elisa Montagni[1*], Xavier Hernando-Momblona[1], Marta Sevillano[1], Alexandre Calon[1], Peter Jung[1], Anna Abuli[3,4], Antoni Castells[3], Sergi Castellvi-Bel[3], Francisco X. Real[2,5] and Eduard Batlle[1,6]

1. Institute for Research in Biomedicine (IRB Barcelona), 08028 Barcelona, Spain
2. Spanish National Cancer Research Centre, 28029 Madrid, Spain
3. Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Catalonia, Spain
4. Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, Barcelona, Catalonia, Spain
5. Universitat Pompeu Fabra, Barcelona, Spain
6. Institució Catalana de Recerca i Estudis Avançats (ICREA).

(*) These authors contributed equally to the work and their names are written in alphabetical order

Correspondence should be addressed to:
Eduard Batlle (eduard.batlle@irbbarcelona.org)

# ABSTRACT

Aberrant activation of WNT signaling and loss of BMP signals represent the two main alterations leading to the initiation of colorectal cancer (CRC). Here we screen for genes required for maintaining the tumor stem cell phenotype and identify the zinc-finger transcription factor GATA6 as key regulator of the WNT and BMP pathways in CRC. GATA6 directly drives the expression of *LGR5* in adenoma stem cells while it restricts BMP signaling to differentiated tumor cells. Genetic deletion of *Gata6* in mouse colon adenomas increases the levels of BMP factors, which signal to block self-renewal of tumor stem cells. In human tumors, GATA6 represses *BMP4* gene expression through binding to a regulatory region that has been previously linked to increased susceptibility to develop CRC. Thus, GATA6 creates a permissive environment for tumor stem cell expansion by controlling the major signaling pathways that influence CRC initiation.

# INTRODUCTION

Constitutive activation of WNT signaling in the intestinal epithelium by mutations in pathway components such as *APC* or beta-catenin leads to the formation of adenomas (Kinzler and Vogelstein, 1996). These precancerous lesions are prevalent in the population and represent the substrate from which colorectal cancer develops (Kinzler and Vogelstein, 1996). An alternative pathway leading to the formation of intestinal tumors is initiated by mutational inactivation of the BMP signaling pathway. Individuals with germline mutations in BMP pathway components such as *BMPRI* or *SMAD4* develop hamartomatous polyps, benign outgrowths of the epithelium that confer high risk of CRC (Brosens et al., 2011; Haramis et al., 2004). In addition, Genome Wide Association Studies (GWAS) of CRC have identified about 30 common low-risk susceptibility variants, four of which map in genomic regions located near the locus encoding for the BMP pathway components *SMAD7*, *GREM1*, *BMP4* and *BMP7* (Broderick et al., 2007; Houlston et al., 2010; Tomlinson et al., 2011b). Recent lineage tracing analysis of *Apc* mutant adenomas in the intestinal epithelium of mice has demonstrated that these tumors are organized according to a stem cell hierarchy that resembles that of normal intestinal crypts (Schepers et al., 2012). The base of adenoma glands hosts a population of Adenoma Stem Cells (AdSCs) that self-renew and give rise to the rest of adenoma cell types. AdSCs express a gene program similar to normal intestinal stem cells (ISCs) whereas non-stem cells are reminiscent of Transit Amplifying (TA) or differentiated cells (Schepers et al., 2012). Genetic ablation of AdSCs in mouse models of intestinal cancer impairs the renewal of the adenoma lineages and induces tumor collapse (Nakanishi et al., 2013). Importantly, several lines of evidence support the notion that full-blown human CRCs rely on a similar stem cell hierarchy (Dalerba et al., 2011; Merlos-Suarez et al., 2011) (Vermeulen et al., 2010).

# RESULTS AND DISCUSSION

The above observations have raised great interest in the biology of intestinal tumor stem cells and their suitability as therapeutic targets. To enable predictions about the mechanisms underlying the tumor stem cell phenotype, we used ARACNe (algorithm for the reconstruction of accurate cellular networks). ARACNe is a method for reverse engineering of gene networks from microarray expression profiles, which identifies statistically significant gene-gene co-regulation by mutual information, an information-theoretic measure of relatedness (Margolin et al., 2006). ARACNe has been previously used to unravel the transcriptional network of several biological systems including B

cells (Basso et al., 2005), malignant glioma (Carro et al., 2010) or T-All leukemia (Della Gatta et al., 2012). We studied associations of genes in CRC by applying ARACNe to a large transcriptomic dataset (n=180 samples) corresponding to primary human colon adenoma, adenocarcinomas and CRC cell lines profiled in different experimental conditions (see methods for details). From the computed genome-wide network, we focused on the ISC gene program (Figure 1A). ARACNe anticipated a subnetwork centered on *LGR5*, a marker gene of both ISCs (Barker et al., 2007) and AdSCs (Schepers et al., 2012). This subnetwork contained many of the best characterized mouse and human intestinal stem cell-enriched genes (orange nodes in Figure1A) including *ASCL2*, *AXIN2*, *EPHB2*, *EPHB3*, *SMOC2* and *ZNRF3* (Dalerba et al., 2011; Jung et al., 2011; Merlos-Suarez et al., 2011; Munoz et al., 2012). These genes are downstream targets of the WNT signaling pathway and are coordinately expressed in CRC-SCs (Dalerba et al., 2011; Jung et al., 2011; Merlos-Suarez et al., 2011; Vermeulen et al., 2010). Additionally, several other nodes (in purple in Figure 1A) have been recently identified as part of the normal ISC expression program in a transcriptomic and proteomic analysis of LGR5+ mouse crypt cells (Munoz et al., 2012). Using commercially available antibodies, we further confirmed that two other nodes that had not been previously connected to ISCs, CLDN2 and SLC12A2, displayed expression patterns restricted to the crypt bottoms and to the base of adenoma glands where normal and tumor stem cells reside, respectively (Supplementary Figure S1; page 59).

Most of the nodes that ARACNe associated to LGR5 corresponded to genes with unknown functions in CRC or in ISCs. We speculated that some of them might represent new regulators of the tumor stem cell phenotype. To explore this possibility, we performed a shRNA screen in LS174T cells, a CRC cell line that retains an ISC-like phenotype (van de Wetering et al., 2002b). We used 4-5 independent shRNAs to target every gene included in the ARACNe ISC subnetwork and subsequently measured *LGR5* gene expression as readout for stemness. Figure 1B displays changes in *LGR5* mRNA levels induced by the two best shRNAs targeting each gene. Supplementary Table S1 (see DVD) details the results of the complete screen. *LGR5* expression in CRC is driven by mutational activation of the beta-catenin/TCF complex (Barker et al., 2007). Fittingly, shRNAs targeting beta-catenin downregulated *LGR5* mRNA in LS174T cells (Figure 1B). In contrast, few of the genes screened were capable of significantly decreasing or increasing baseline *LGR5* gene expression (Figure 1B). Notably, knockdown of the zinc-finger transcription factor *GATA6* decreased LGR5 mRNA levels to a similar extent to that triggered by beta-catenin knockdown (Figure 1B).

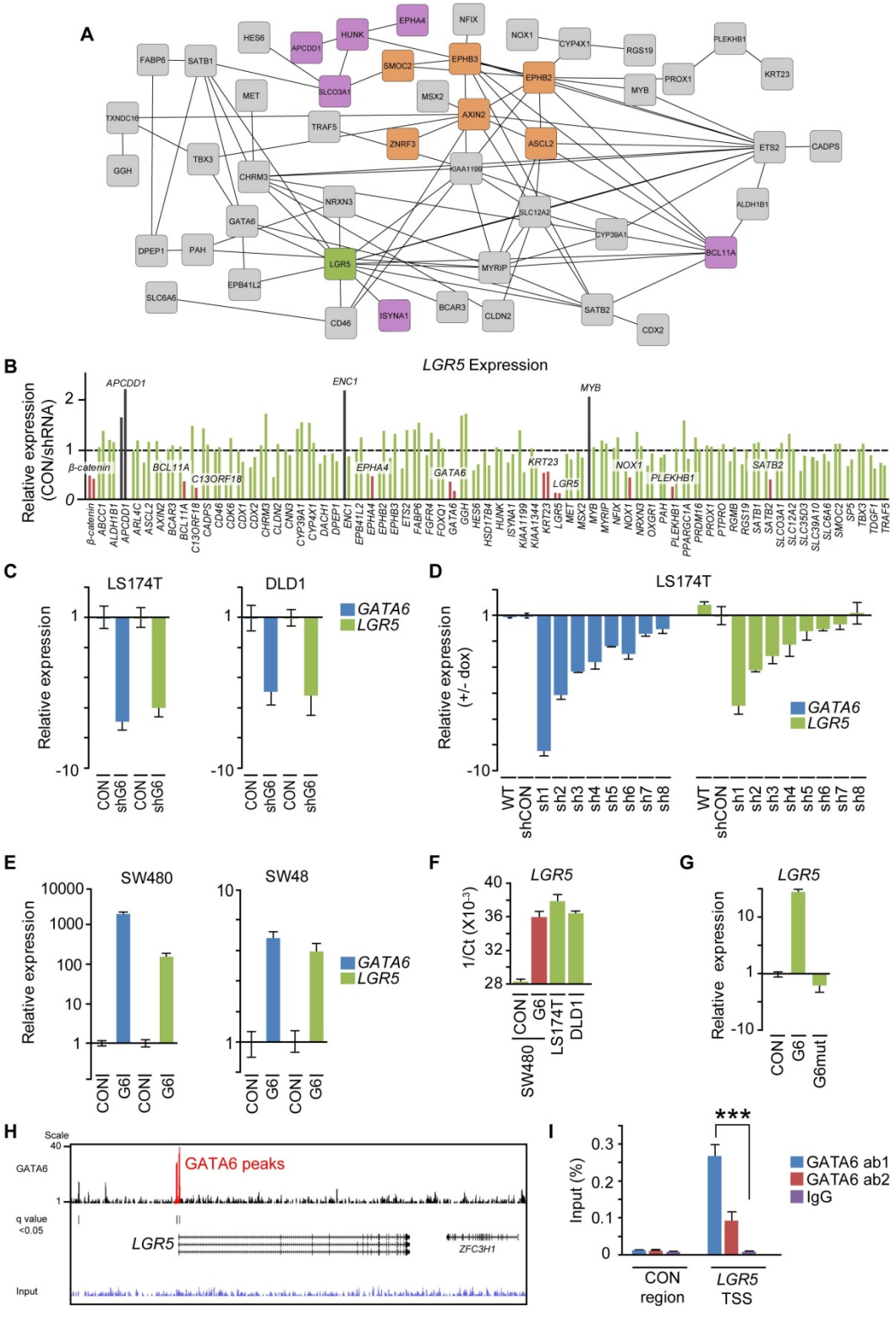



**Figure 1. The transcription factor GATA6 is a regulator of LGR5 expression.** (A) ARACNe network of genes transcriptionally connected to the bona-fide stem cell marker *LGR5* (green) and other intestinal stem cell (ISC) markers (purple and orange). All other genes within the network are depicted in grey. (B) shRNA screen results based on ARACNe prediction. Relative gene expression levels of *LGR5*, determined by qRT-PCR in LS174T cells transduced with the two best shRNAs for each gene in the ARACNe network are shown (see Table S1 (see DVD)

for more details on all shRNA used). Significant up- or downregulation of *LGR5* expression for the respective shRNAs are shown in black or red bars, respectively. shRNAs that did not significantly affect *LGR5* expression are shown in green. (C) Relative gene expression of *GATA6* (blue bars) and *LGR5* (green bars) in LS174T (left) or DLD1 (right) cells stably transduced with control shRNA (CON) compared to cells stably expressing a shRNA against *GATA6* (shG6). Data are shown as mean±s.e.m. of three independent experiments. (D) To confirm the specificity of the *GATA6* shRNA, the two best and six additional shRNAs were cloned into a Doxycycline inducible shRNA system. Relative gene expression of *GATA6* (blue bars) and *LGR5* (green bars) in LS174T wild type cells (WT) and cells stably transduced with control shRNA (shCON) compared to cells stably expressing short hairpins against *GATA6* (sh1-8) are shown. Data are mean±s.e.m. of three independent experiments. (E) Gene expression of *GATA6* (blue bars) and *LGR5* (green bars) in SW480 and SW48 cells overexpressing *GATA6 (*G6) compared to control cells (CON). Data are shown as mean±s.e.m. of three independent experiments. (F) Overexpression of *GATA6* in SW480 (SW480 G6, red bar) compared to cells transduced with an empty vector (CON) (green bar), restores *LGR5* gene expression to levels typically observed in colorectal cancer cell lines LS174T and DLD1 (green bars). (G) Regulation of *LGR5* expression by GATA6 requires its DNA binding domain**.** SW480 cells were transduced with an empty vector (CON), *GATA6* (G6) or *GATA6* with point mutations in both zinc fingers (G6mut) and expression of *LGR5* (green) was determined by q-PCR. Data are mean±s.e.m. of three independent experiments. (H) GATA6 ChIP-seq peaks (red) in *LGR5* promoter region, displayed in the UCSC genome browser. Input is shown in blue. ChIP-seq was performed using 3 biological replicates and peaks are summarized into a single track (black). Significant GATA6 peaks were identified using a q value <0.05. (I) GATA6 binds to the *LGR5* promoter. ChIP PCR in LS174T using two independent antibodies against GATA6 (blue and red bars) and IgG control (purple bars). Data are represented as % of input chromatin, compared with non-specific IgG (purple) binding to a genomic control region. Data are shown as mean±s.e.m. of four independent experiments; Student t-test, ***P<0.01.

The above result prompted us to investigate the role of GATA6 in the regulation of LGR5. GATA6 knockdown induced silencing of LGR5 in both LS174T and DLD1 CRC cells (Figure 1C). The extent of *GATA6* knockdown mirrored that of *LGR5* downregulation as shown by 8 different doxycycline inducible shRNAs (Figure 1D). The CRC cell lines SW480 and SW48 possessed nearly undetectable or very low expression of both *GATA6* and *LGR5* compared with other CRC cells. Re-expression of *GATA6* in these two cell lines raised *LGR5* mRNA levels to those present in LS174T and DLD1 CRC cells (Figure 1E-F and data not shown). A point mutation in the zinc-finger domain that impairs binding of GATA6 to DNA abrogated these effects (Figure 1G). Finally, chromatin immunoprecipitation using anti-GATA6 antibodies followed by massively parallel DNA sequencing (ChIP-seq) revealed prominent GATA6 binding to the *LGR5* proximal promoter in LS174T cells (Figure 1H). We could further demonstrate a large enrichment in GATA6 bound to this region by ChIP using two independent antibodies (Figure 1I). Overall, these results demonstrate that GATA6 directly regulates the expression of the *LGR5* locus in CRC cell lines.

The intestinal epithelium expresses two GATA factor homologues, GATA4 and GATA6. We found that GATA6 was expressed homogenously in mouse small intestine and colon adenomas as well as in human CRCs whereas GATA4 was

restricted to small intestine (Supplementary Figure S2A; page60). We thus focused on the effects of GATA6 in colonic tumorigenesis. To this end, we generated mice carrying conditional knock-out alleles of the tumor suppressor gene *Apc* and *Gata6* in ISCs by means of the *Lgr5*-eGFP-CreERT2 driver (Barker et al., 2009). To increase tumor burden in the colorectum, we combined deletion of the floxed alleles by tamoxifen administration with DSS treatment (Tanaka, 2012) (see methods). This protocol led to death of all *Apc* mutant mice within 8 months due to generation of multiple adenomas throughout the intestinal tract (median survival=96 days; Figure 2A-B). Mortality was largely reduced in mice with compound *Gata6* and *Apc* deletions (median survival=223 days) (Figure 2A). Macroscopic inspection of the intestines revealed reduced extension of adenomatous tissue in $Lgr5^{eGFPCreERT2}Gata6^{fl/fl}Apc^{fl/fl}$ compared to $Lgr5^{eGFPCreERT2}Gata6^{+/+}Apc^{fl/fl}$ intestines (Figure 2B and quantification in Figure 2C). Histological examination demonstrated that adenomas arising in the colon of $Lgr5^{eGFPCreERT2}Gata6^{fl/fl}Apc^{fl/fl}$ mice lacked GATA6 expression (examples in Fig 2D-E). Adenoma Stem Cells (AdSC) could be readily visualized by the *Lgr5*-GFP reporter (Schepers et al., 2012) (arrowheads in Figure 2H-I). We observed a sharp decrease in the frequency of *Lgr5*-GFP-high tumor cells in *Gata6* mutant adenomas both by immunohistochemistry (Figure 2H-I and examples in Supplementary Figures S3 and S4; pages 61-62) and by flow cytometry of dissected adenomatous tissues (Figure 2J. n=6 mice). It was apparent that the localization of *Lgr5*-GFP+ cells in *Gata6* deficient tumors was further restricted to the bottommost positions of adenoma glands, in proximity to submucosal layers (Figure 2I and examples in Supplementary Figures S3 and S4; pages 61-62). Cldn2+ and Slc12a2+ tumor cells also displayed reduced frequencies in *Gata6* mutant tumors (Supplementary Figures S3 and S4; pages 61-62).

We also acutely deleted *Apc* and *Gata6* throughout the adult intestinal epithelium by breeding the above strains to *Villin*-CreERT2 mice (Supplementary Figure S5; page 63). Complete *Apc* deletion in the intestinal epithelium caused death after 1 week due to failure of intestinal function, which precludes long-term analysis. Nevertheless, the intestinal epithelium of $Villin^{CreERT2}Gata6^{fl/fl}Apc^{fl/fl}$ mice showed reduced proliferation compared to $Villin^{CreERT2}Gata6^{+/+}Apc^{fl/fl}$ littermates (Supplementary Figure S5A; page 63). We confirmed decreased frequency of AdSCs in $Villin^{CreERT2}Gata6^{fl/fl}Apc^{fl/fl}$ mice by introducing the *Lgr5*-GFP reporter allele or by CLDN2 IHC (Supplementary Figure S5B-C; page 63). Of note, the decreased tumorigenicity and reduced AdSC numbers induced by *Gata6* deficiency was less prominent in the small intestine than in the colon of both *Villin*-CreERT2 and *Lgr5*-CreERT2 driven models (Supplementary Figure S6; page 64). This result is compatible with a redundant role of GATA4 and GATA6 in the small intestine (Supplementary Figure S2; page 60). Indeed, GATA4 and GATA6 bind the same core DNA

consensus sequence (Sakai et al., 1998) and we found that both were capable of reconstituting *LGR5* and *CLDN2* gene expression in SW480 cells (Supplementary Figure S2B; page 60).

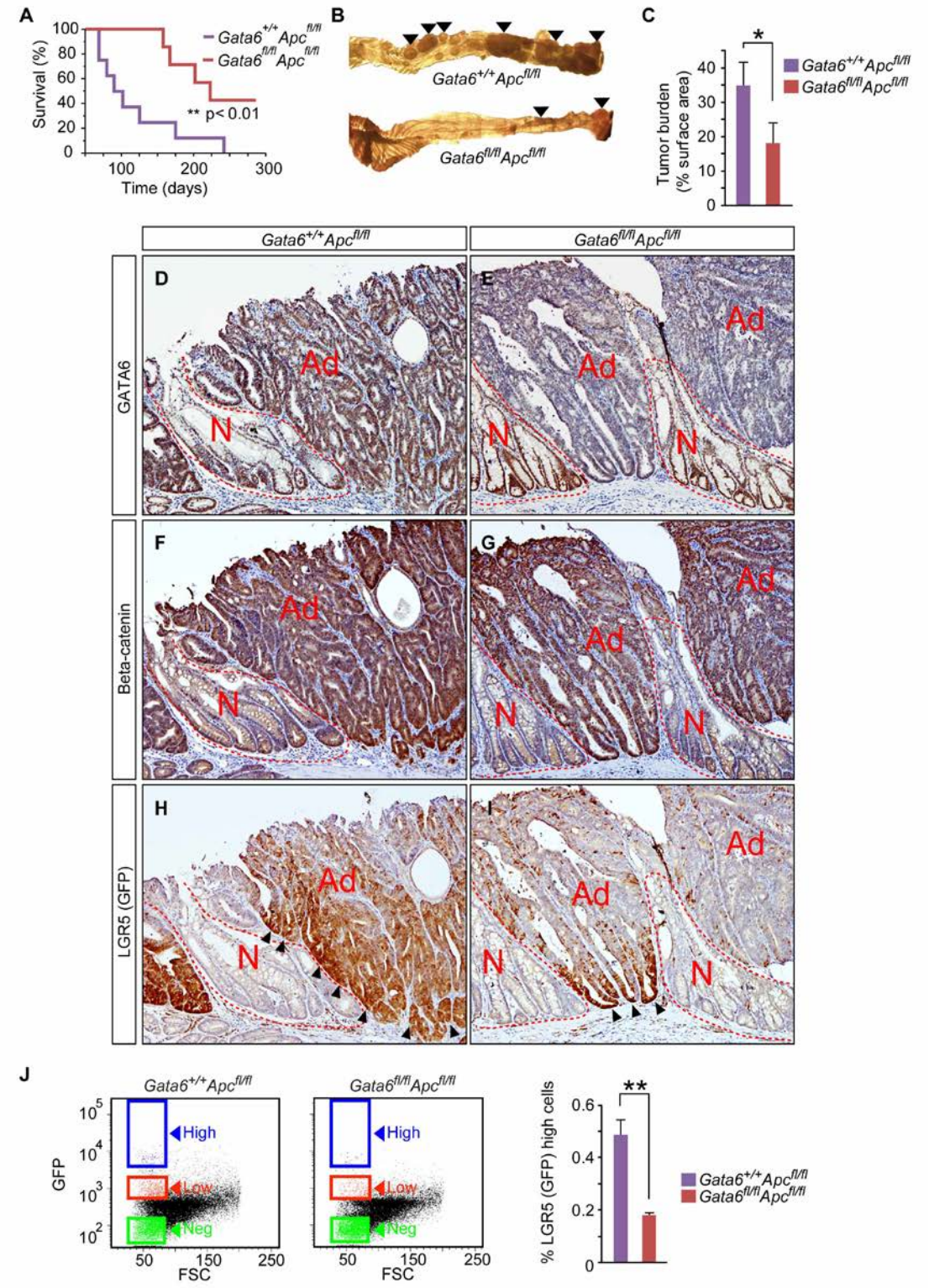


**Figure 2. *Gata6* deletion in an *Apc* null background leads to prolonged survival, lower tumor burden and a decrease in LGR5 positive cells in colon adenomas. (**A) Survival analysis indicated significantly increased survival probability of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red line) compared to *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple line) mice. Log rank test, ** P<0.01, n=8 mice/genotype. (B) Representative macroscopic images of the colon from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ animals treated with Tamoxifen and DSS in the colon (see methods for details). (C) Tumor burden in the colon in the above cohort of mice. Median tumor surface area was significantly lower in the *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bar) compared to *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ *(*purple bar). Data are shown as mean±s.e.m. of five mice/genotype; Student t-test, *P<0.05. (D-I) Decreased LGR5 expression in *Gata6* null colon adenomas. Representative murine colon adenomas stained for GATA6 (D-E), Beta-catenin (F-G) and LGR5 (GFP) (H-I) from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) animals treated with 2% DSS. Normal (N) tissue is delimited by a red dashed line from adenoma (Ad) tissue. Adenoma tissue can be readily distinguished by beta-catenin accumulation. Black arrowheads point to LGR5 (GFP-high) positive cells located at the bottom of adenoma glands. (J) Colonic tumors from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) mice were disaggregated and analyzed by flow cytometry. GFP-high (blue), -low (red) and -negative cells (green) were gated according to log GFP fluorescence intensity as described elsewhere[7]. Quantification of % LGR5 (GFP) high tumor cells from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice. Data are shown as mean±s.e.m. of six mice; Student t-test, **P<0.01.

To further analyze the function of GATA6 in tumors, we cultured adenomas that arose in the colon of wild-type and mutant mice (Figure 3A-C). Purified *Apc* mutant Lgr5-GFP$^{high}$ cells expand as 3D tumor organoids in conditions equivalent to those required for propagating normal ISCs with the exception that they do not require WNT factors (i.e. Wnt3a or R-Spondin) or BMP antagonists (i.e. Noggin). As described elsewhere (Sato et al., 2010), they formed hollow spheroid structures that expand indefinitely in culture (Figure 3A, left and 3B). In contrast, GFP$^{high}$ cells from *Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ formed tumor organoids displaying multiple crypt-like folds (Figure 3A, right) that grew with slow kinetics (Figure 3B). Previous work had shown that the capacity to expand in vitro organoids from mouse adenomas is largely restricted to the Lgr5-GFP$^{high}$ cell population implying that *in vitro* clonogenic capacity is a surrogate of stemness (Schepers et al., 2012). We thus assessed the functionality of AdSCs by measuring tumor organoid forming capacity upon serial passaging (Figure 3C). These experiments revealed that *Gata6* deficient adenomas progressively lost clonogenic cells. As a result, we failed to propagate *Gata6* null tumor organoids beyond the fourth week (Figure 3C). These observations indicate that *Gata6* deficiency impairs AdSC self-renewal.

To understand the requirement of *Gata6* for AdSC function, we profiled the expression of *Gata6* wild-type and mutant tumor organoids (Supplementary Table S2 (see DVD)). Gene set enrichment analysis (GSEA) indicated that WNT signaling components were upregulated upon *Gata6* deletion (Supplementary Table S3 (see DVD)). In particular, we observed a massive

(>100 fold) and coordinated upregulation of many well-established negative regulators of the WNT pathway such as *Apcdd1*, *Dkk3*, *Nkd1*, *Notum*, *Wif1* and *Znrf3* (Supplementary Figure S7; page 65). All these molecules inhibit WNT signaling upstream of *Apc* and we could not find evident alterations in nuclear beta-catenin accumulation in *Gata6* mutant adenomas (examples in Figure 2F-G and Supplementary Figures S4 and S5; pages 64-65). Thus, it is unclear whether this anti-WNT response could have a negative impact on the formation of adenomas in our experimental setting. In addition, GSEA revealed a prominent and significant enrichment in components of the TGF-beta superfamily most of which belong to the BMP signaling pathway (Supplementary Table S3 (see DVD)). We confirmed that *Gata6* deficiency induced the upregulation of *Bmp2*, *4* and *7* as well as of bona-fide BMP target genes such as *Id1*, *Id3* and *Smad6* (Figure 3D). Consistent with this finding, *Gata6* mutant tumor organoids displayed accumulation of nuclear P-SMAD1/5/8 owing to BMP pathway activation (Figure 3E).

The above results prompted us to investigate the role of BMP signaling in the expansion of AdSCs mediated by GATA6. Deletion of *Apc* in the mouse intestinal epithelium increased P-SMAD1/5/8 positivity compared to normal colon (Figure 3F). Therefore, mutational activation of the WNT pathway *in vivo* concurs with increased BMP signaling. Remarkably, we noticed that P-SMAD1/5/8 positivity was heterogeneous in adenomas. Stronger staining corresponded to the upper regions of adenomatous crypts (Figure 3F and 3F' – arrows) whereas the bases of tumor glands were weakly stained (Figure 3F and 3F' - arrowheads). These areas hosted Lgr5-GFP-high cells (serial sections in Figure 3H and 3H' - arrowheads) implying relatively lower BMP signals in AdSCs. Compound deletion of *Apc* and *Gata6* exacerbated BMP signaling in the intestinal epithelium and extended the P-SMAD1/5/8 positive domain into the AdSC compartment (Figure 3G and 3G' –arrows). Consistent with this observation, the BMP target ID1 and LGR5 displayed complementary expression patterns in adenomas arising in *Gata6* wild-type littermates (Figure 3J). AdSCs located at the bottommost positions of tumor glands were ID1 negative (Figure 3J, white arrowheads) whereas ID1 expression marked differentiated tumor cells localized in upper regions (Figure 3J, arrows) as well as a subset of LGR5+ cells at intermediate positions in tumor glands (Figure 3J. yellow arrowheads). In contrast, the ID1+ domain reached the bottoms of the glands in *Gata6* null adenomas (Figure 3K, arrows). As a result, all Lgr5-GFP+ cells gained Id1 expression (Figure 3K, arrows). We confirmed extended P-SMAD1/5/8+ domain and expansion of ID1+ domain into the base of tumor glands upon acute deletion of *Apc* and *Gata6* in the whole epithelium using the *Villin*-CreERT2 driver (Supplementary Figure S8; page 66).

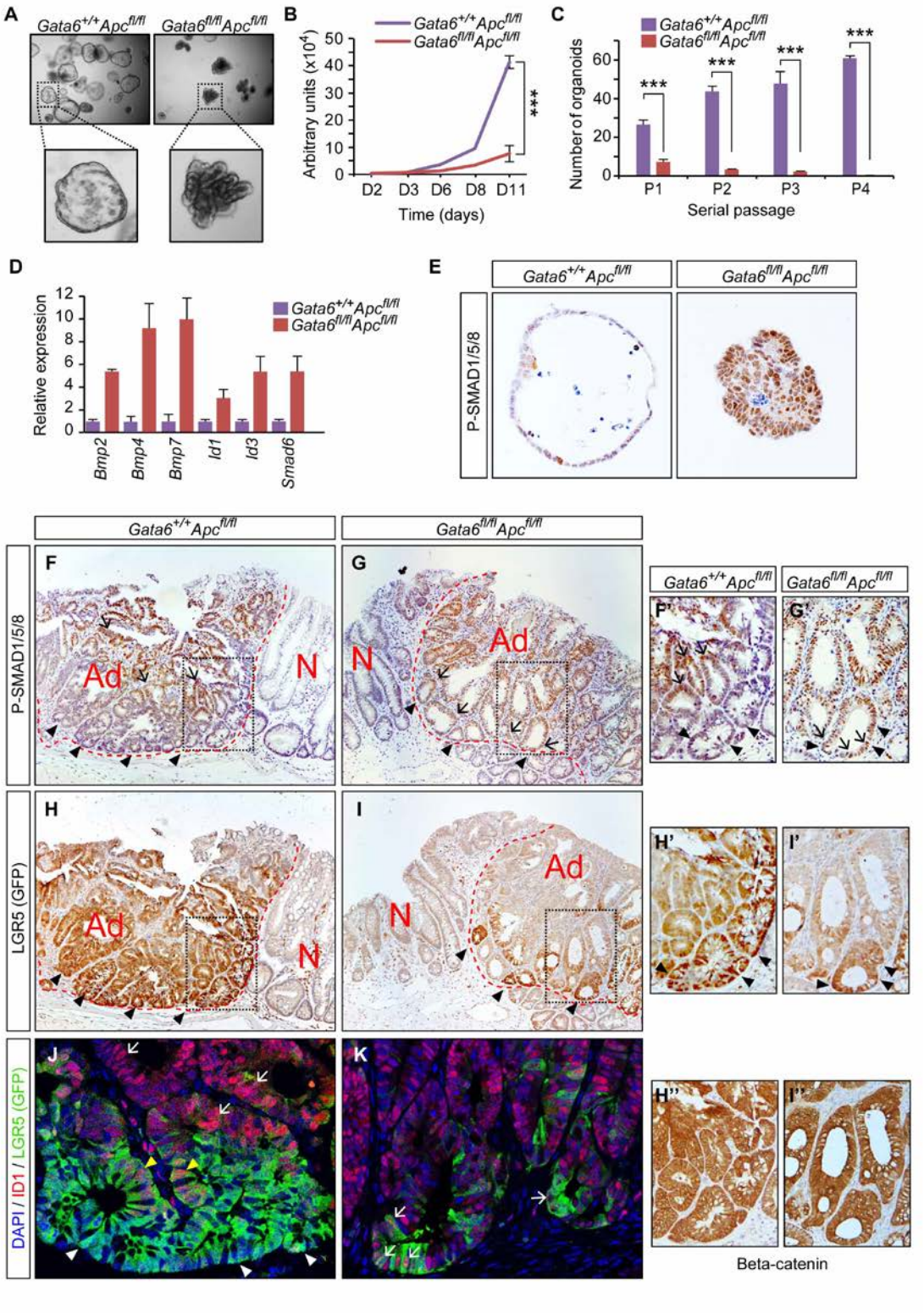



**Figure 3. *Gata6* deletion in colon adenoma cells results in increased BMP signaling. (**A) Morphology of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ tumor organoids cultured under standard conditions for 7 days. (B) *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red) tumor organoid cultures were assessed for their growth potential. Data are mean±s.e.m. of four replicates, Student t-test, ***P<0.001. (C) Serial passaging of *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bars) tumor organoid

cultures. Note the loss of clonogenicity in compound mutants. (D) *Gata6* null tumor organoids (red bars) aberrantly activate BMP pathway components compared to *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) as determined by qRT-PCR. Data are mean±s.e.m. of three independent experiments. (E) Representative P-SMAD1/5/8 staining of colon tumor organoids from *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) tumor organoids cultured under standard conditions. (F-I) Representative serial stainings of P-SMAD1/5/8 (F-G) and LGR5 (GFP) (H-I) of colonic adenomas from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right). Red dashed lines delineate normal (N) and adenoma (Ad) tissue. AdSCs are readily identified by LGR5 (GFP) staining (black arrowheads). High P-SMAD1/5/8 staining (blue arrows) is localized to the upper portion of *Apc* null tumors, while this positivity extends into the AdSC zone in compound mutants. (F'-I'') Insets of SMAD1/5/8 (F',G'), LGR5 (GFP) (H',I'), and Beta-catenin (H'', I'') staining of tumors from F-I. Note that in *Gata6* null tumors the P-SMAD1/5/8 gradient extends into the bottommost position of the tumor glands where LGR5 GFP-high cells reside. (J-K) Immunofluorescence stainings of ID1 (red) and LGR5 (GFP) (green). Nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI, in blue). White arrowheads point to the AdSC zone, yellow arrowheads point to ID1/LGR5-GFP double positive cells at intermediate positions and arrows to differentiated cells that express ID1 positioned in the upper part of adenomatous glands. Note that the ID1 domain extends well within the stem cell zone in *Gata6* null adenomas (arrows).

We next explored whether increased BMP signaling was causal in the impairment in self-renewal capacity of *Gata6* deficient AdSCs (Figure 4). Treatment of *Gata6* wild-type organoids with recombinant BMP2 or BMP4 reproduced most of the features of *Gata6* deficiency such as wrinkled morphology (Figure 4A) and elevated expression of WNT inhibitors (Figure 4B).

Remarkably, *Gata6* mutant tumor organoids treated with the BMPR1 specific inhibitor DMH1 (Hao et al., 2010) or with the BMP antagonist Noggin (data not shown) displayed the hollow morphology of wild-type tumor organoids and recovered growth rates (Figure 4A and 4C). DMH1 treatment also decreased expression of the negative regulators of the WNT pathway triggered by *Gata6* deletion (Figure 4B). Analysis of clonogenic potential confirmed that DMH1 rescued the self-renewal ability of *Gata6* mutant AdSCs (Figure 4D).
Indeed, continuous treatment with DMH1 allowed long-term propagation of *Gata6* mutant tumor organoids (>6 months) with efficiency similar to that of their wild-type counterparts (data not shown).

To complete this study, we investigated the regulation of BMP genes by GATA6 in human tumors. In adenomas from patients with Familial Adenomatous Polyposis (FAP, n=5 independent samples) prominent P-SMAD1/5/8 staining mirrored nuclear beta-catenin accumulation (Figure 5A, and examples in Supplementary Figure S9; page 67). Similarly to mouse adenomas, *BMP4* was upregulated upon *GATA6* knockdown in human LS174T CRC cells (Figure 5B). Previous work had demonstrated that beta-catenin/TCF4 activity positively controls *BMP4* levels in CRC cells (van de Wetering et al., 2002b). Consistent with this finding, we showed that a tamoxifen-inducible dominant negative TCF4

(N-Terminus TCF4 fused to ERT2) that disrupts beta-catenin/TCF4 interaction repressed BMP4 expression in LS174T cells (Figure 5C). We thus mapped GATA6 occupancy in LS174T cells by ChIP-seq and compared it with that of the beta-catenin/TCF4 complex (Figure 5D). These experiments revealed no GATA6 bound to the *BMP4* proximal promoter. Instead, three regions located -20 Kb (R1), -50 Kb (R2) and +120kb (R3) from the transcription start site displayed significant enrichment in GATA6 binding (Figure 1D). Data collected for the ENCODE project in a panel of cell lines indicated that the 3 regions bound by GATA6 overlap to a large extent with DNAse I hypersensitive sites (Figure 5D). They were also enriched in H3K4me1 (Figure 5D), a histone modification that marks enhancer regions (Heintzman et al., 2009). The nearest coding region to these three GATA6 binding sites was *BMP4* whereas the expression of the *BMP4* flanking genes (*CDKN3* and *DDH1*) remained unaltered by *GATA6* knockdown (data not shown). Therefore, R1, R2 and R3 may likely represent enhancers of the *BMP4* locus. Interestingly, we noticed that a SNP (rs1957636) located about 10Kb from R3 (Figure 5E) had been previously linked to CRC susceptibility in GWAS studies (Tomlinson et al., 2011b). We further inspected linkage disequilibrium (LD) blocks at this genomic location within the 1000 Genomes Project data (Abecasis et al., 2010) using Haploview v4.2(Barrett et al., 2005) in the European population (CEU). This analysis identified three additional SNPs (rs1951674, rs713424 and rs728425) located within GATA6 binding regions at R3 that appear to be in linkage disequilibrium (each at $r^2$=0.688, D'=1.00) with rs1957636 (Figure 5E). We confirmed specific binding of GATA6 to several peaks within R3 by ChIP followed by qPCR (Figure 5F). Previously published ChIP-seq data of beta-catenin and TCF4 occupancy in LS174T cells indicate that beta-catenin/TCF4 complex binds to R3.1 (Figure 5E, red box) (Mokry et al., 2010; Mokry et al., 2012). We validated occupancy of R3.1 by beta-catenin and TCF4 in LS174T cells and demonstrated that knockdown of GATA6 facilitated binding of these downstream effectors of the WNT pathway to this region (Figure 5G-H). We also found increased levels of the transcriptional co-activator p300 bound to R3.1 upon *GATA6* knockdown (Figure 5I). These observations suggest a direct repressive activity of GATA6 on the *BMP4* R3.1 enhancer in human CRC.

Overall, our data reveals the existence of a transcriptional circuit dedicated to suppress the renewal of tumor stem cells through BMP signaling. This circuit is switched on coinciding with the over-activation of the WNT pathway in adenomas and it is negatively regulated by GATA6. Previous works have shown that genetic inhibition of the BMP pathway in the intestinal epithelium of mice specifies the formation of extra stem cell niches in the differentiated compartment leading to overgrowths of the epithelium (Haramis et al., 2004).

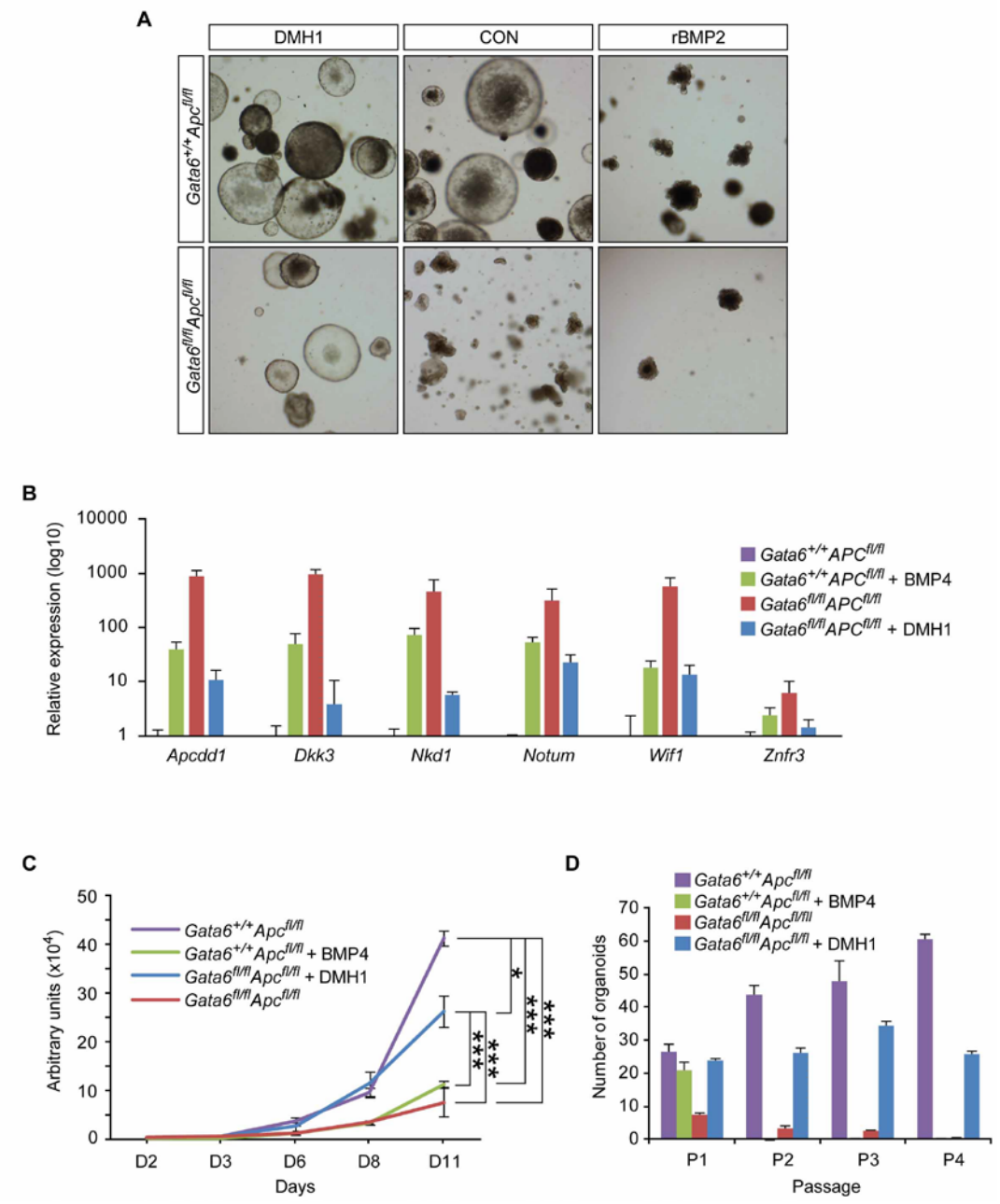



**Figure 4. BMP inhibitors rescue morphological, proliferative and clonogenic potential of *Gata6* null tumor organoids. (**A) Colon tumor organoids from *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice cultured in control media (middle) or media supplemented with rBMP2 (right) or BMP pathway antagonist DMH1 (left) for 12 days. Note that DMH1 restores spheroid shape in compound mutants. (B) Compound mutant tumor organoids highly upregulate expression of WNT inhibitors. *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red) organoid cultures were cultured in the presence of either rBMP4 (green) or DMH1 (blue). Note that treatment of these tumor organoids with BMP inhibitor DMH1 decreases WNT inhibitor upregulation. Data are mean±s.e.m. of three independent experiments. (C) Tumor organoids were cultured as in B and assessed for their growth potential. Compound mutants grown in the presence of DMH1 recover their proliferative potential, whereas rBMP4 treatment significantly reduces proliferation in *Apc* null tumor organoids. Data are mean±s.e.m. of four replicates; Student t-test, *P<0.05, and ***P<0.001. (D) Serial passaging of organoid cultures under the same conditions as in B. Note the loss of clonogenicity in BMP4 treated (green) and compound mutants (red). DMH1 treatment rescues clonogenicity (blue).

On the contrary, enforcement of BMP4 signaling blocks the tumor initiating potential of late stage CRC-SCs (Lombardo et al., 2011). The mechanism by which BMP impairs tumor stem cell expansion is not fully understood although inhibition of PI3K/AKT signaling has been proposed (Lombardo et al., 2011). Here, we show that in colon adenomas, BMP signaling induces the coordinated expression of several negative regulators of the WNT pathway. It remains to be elucidated whether any of these molecules can suppress mutational activation of the beta-catenin/TCF complex. We have not investigated whether the GATA6-BMP transcriptional circuit also operates in the normal intestinal epithelium. However, mice with an intestinal epithelial deletion of *Gata6* display decreased crypt proliferation but no changes in WNT target gene levels including *Lgr5* (Beuling et al., 2012; Beuling et al., 2011).

Adding to the notion that colorectal adenomas are built on a stem cell hierarchy, we reveal that these tumors contain two compartments defined by low and high BMP signaling. AdSCs localize outside the BMP signaling positive areas at the base of adenoma glands. This observation is compatible with the existence of a niche at this position that protects tumor stem cells from BMPs. Deletion of *Gata6* increases BMP levels which likely overrides the BMP inhibitory signals in AdSC and blocks self-renewal. Consistent with this model, several negative regulators of the BMP pathway are secreted by stromal cells in proximity to the base of normal and tumor glands (Kosinski et al., 2007). One such BMP inhibitor, Gremlin1 (GREM1), is expressed by subepithelial fibroblasts at the base of the crypts (Kosinski et al., 2007). Interestingly, individuals bearing genomic duplication in a regulatory region of the *GREM1* locus display constitutive upregulation of GREM1 levels in the intestinal epithelium and are at high risk of developing colon polyps and CRC(Jaeger et al., 2012). We speculate that this particular genomic alteration may act by increasing the size of the BMP negative AdSC niche, thus facilitating adenoma expansion at the onset of tumorigenesis. We also show that GATA6 defines a set of enhancers at the human *BMP4* locus near a SNP that has been associated to CRC susceptibility (rs1957636). GATA6 blocks binding of beta-catenin/TCF4 complex to these regulatory regions. It was already reported that rs1957636 does not map to a region of functional relevance and that other SNPs in that region were likely causative of the association with CRC (Tomlinson et al., 2011b). We identify three SNPs in linkage disequilibrium with rs1957636 at GATA6 bound regions. Sequence analysis indicated that these SNPs do not modify GATA6 core binding sequences themselves (data not shown). Thus, future studies should aim to identify additional transcription factors that, acting in combination with GATA6 on *BMP4* enhancers, could be directly affected by the haplotypes that have been associated to CRC risk.

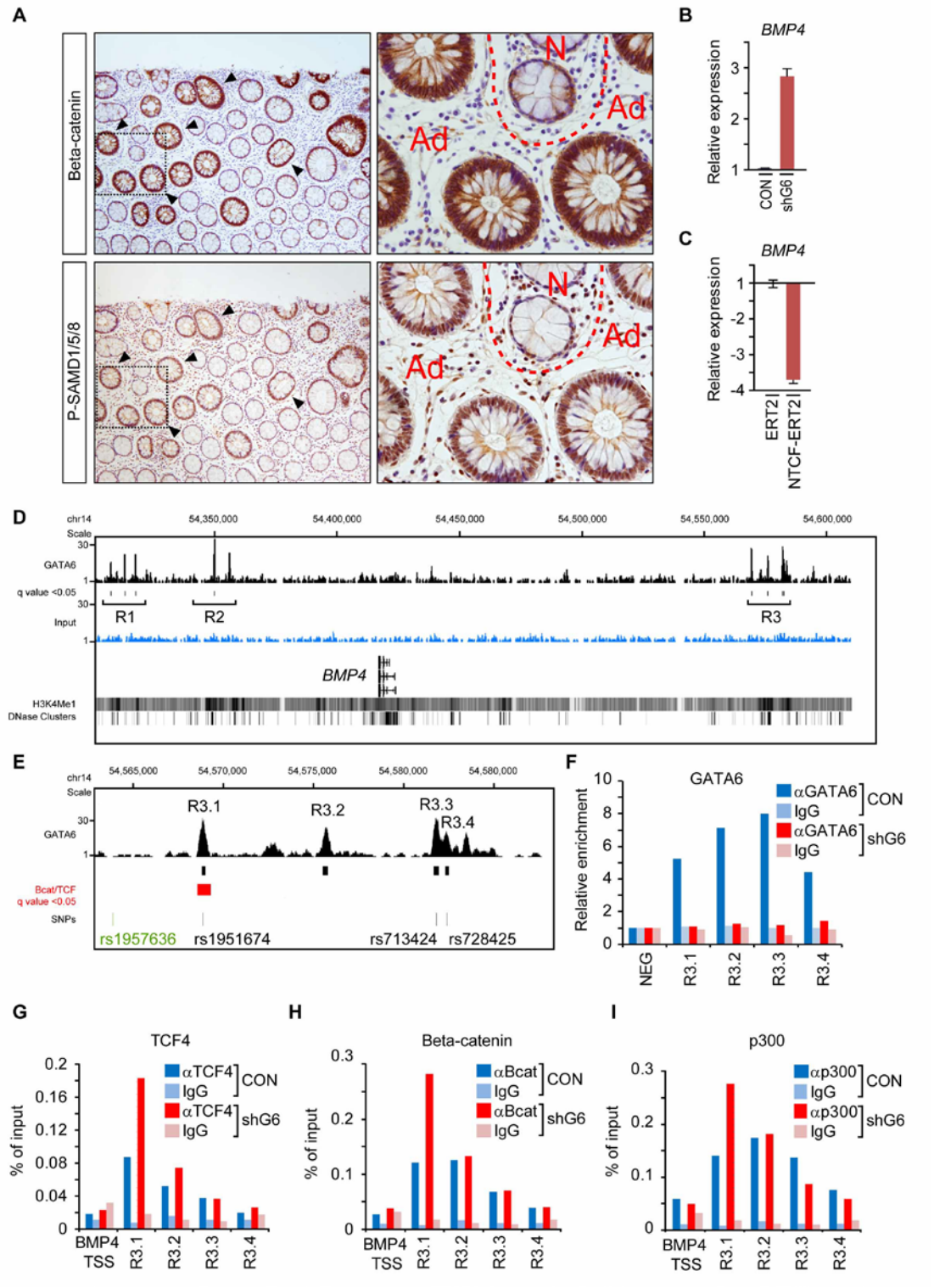



**Figure 5. GATA6 competes with beta-catenin/TCF4 for binding at BMP4 enhancers. (**A) Representative stainings of beta-catenin (top) and P-SMAD1/5/8 (bottom) in serial sections of human adenoma from an individual suffering from FAP. Note nuclear accumulation of beta-catenin and prominent accumulation of P-SMAD1/5/8 in adenoma tissue. Red dashed lines delineate normal (N) and adenoma (Ad) tissue. Insets (black dashed boxes, left) of beta-catenin (top, right) and P-SMAD1/5/8 (bottom, right). **(**B) Relative gene expression of *BMP4* as

determined by qRT-PCR in LS174T cells expressing a doxycycline inducible non-targeting control shRNA or shRNA targeting *Gata6* (both in the presence of doxycycline). (C) *BMP4* levels as determined by qRT-PCR in LS174T cells stably transduced with a 4-hydoxytamoxifen (4OHT) inducible N-terminus domain of TCF4 (NTCF) fused to the modified ligand binding domain of the estrogen receptor (ERT2) compared to cells transduced with the ERT2 alone, both in the presence of 4OHT. Data are mean±s.e.m. of three independent experiments (D) GATA6 occupancy of the *BMP4* locus by ChIP-seq. Three distal regions (R1, R2 and R3) display significant GATA6 binding (q<0.05). ENCODE data for H3K4Me1 and DNAse-I hypersensitive sites is overlaid (E) GATA6 occupancy of *BMP4* R3 with beta-catenin and TCF4 occupancy (see refs 28 and 29) is overlaid. A single region was significantly enriched (q<0.05) in beta-catenin and TCF4 binding (in red), which coincided with GATA6 peak R3.1. The position of the SNPs displaying linkage disequilibrium with rs1957636 (Tomlinson et al., 2011b) are also shown. (F) ChIP of LS174T cells followed by PCR to amplify R1, R2 and R3. Cells were stably transduced with a doxycycline inducible *shGATA6* and were treated (shG6) or not treated (CON) with doxycycline for 4 days. Chromatin was either immunoprecipitated using anti-GATA6 or non-specific IgGs of the same species and isotype. Data are represented as relative enrichment compared to an unrelated genomic control region of replicates, representative values are shown. (G-I) *GATA6* knockdown leads to increased binding of beta-catenin, TCF4 and p300 to the R3.1 BMP enhancer. ChIP was performed under the same conditions as in F with antibodies against TCF4, beta-catenin, p300 or non-specific IgGs of the same species and isotypes. ChIP followed by PCR confirmed that *GATA6* knockdown (red bars) lead to increased occupancy of TCF4, beta-catenin and p300 at the *BMP4* enhancer (R3.1). Data are represented as % of input chromatin of replicates, representative values are shown.

# ACKNOWLEDGMENTS

We thank Elena Sancho for crucial assistance and manuscript proofreading, Hans Clevers and Jurian Schuijers for help interpreting the beta-catenin/TCF4 ChIP-seq data, IRB Biostatistics Unit for analysis of the genomics and transcriptomics data, IRB Functional Genomics Core Facility for technical assistance in microarray hybridization experiments, Begoña Canovas for assistance with histology and the Batlle laboratory for discussions and support. This work has been supported by grants to E.B. from the European Research Council (FP7-EU) and Consolider programs (MICINN).

# METHODS

### Mice

Mouse experimentation protocols were approved by the animal care and use committee of the Barcelona Science Park (CEEA-PCB) and the Catalan Government (P18-R5-09). The mouse lines used in this work were the intestinal stem cell specific Cre (*Lgr5*$^{eGFPCreERT2}$) (Barker et al., 2007) (kindly provided by H. Clevers), the intestinal epithelium specific Cre (*Villin*$^{CreERT2}$) (el Marjou et al., 2004) (kindly provided by S. Robine), conditional Apc$^{fl/fl}$ (kindly provided by

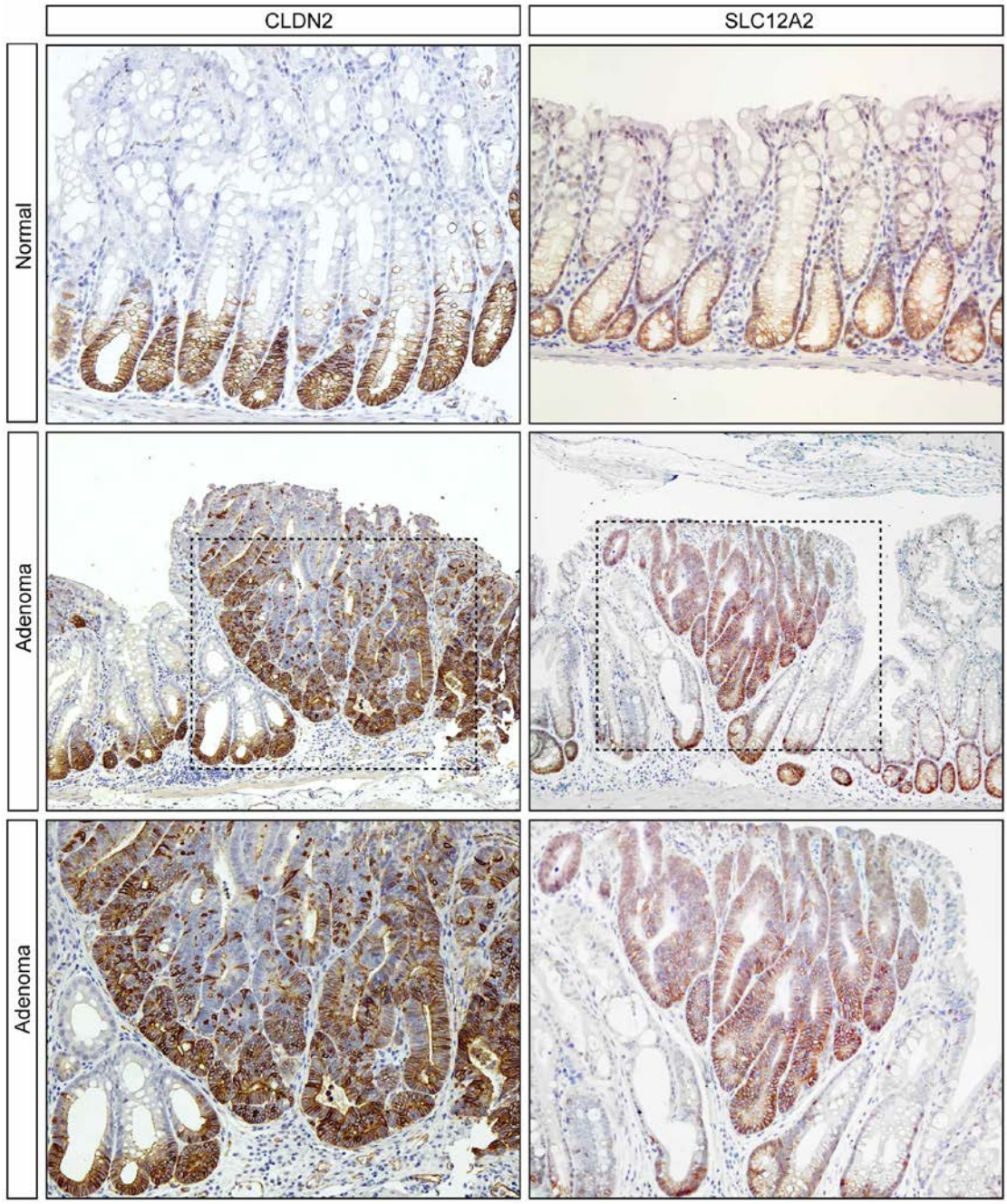
Christine Perret)(Colnot et al., 2004) and conditional *Gata6* (*Gata6*$^{fl/fl}$) (Sodhi et al., 2006) purchased from Jackson Laboratories. All experiments used littermate controls established in a pure C57BL/6 background.

**Tumor induction**

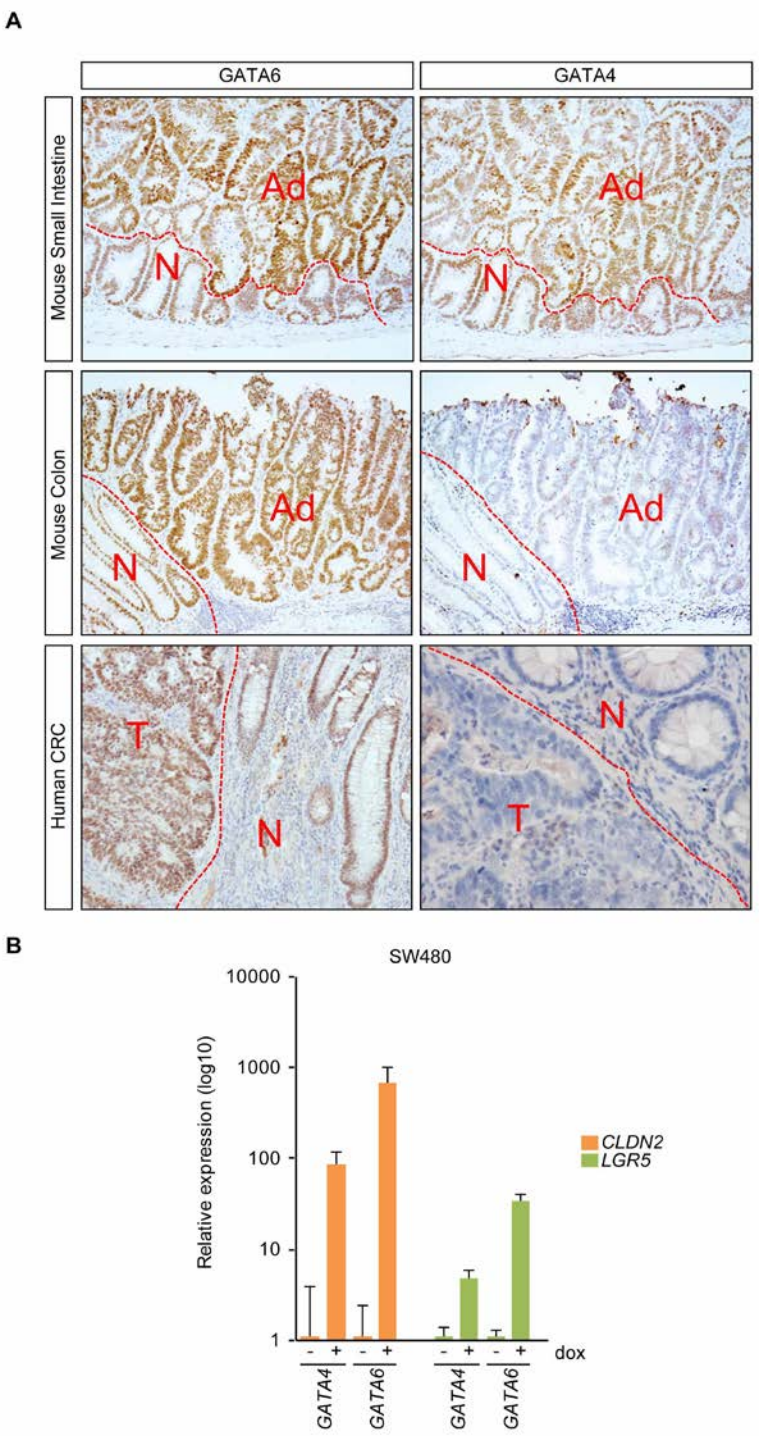
Full recombination using the *Villin*$^{CreERT2}$ driver was induced in control and experiment mice by daily intraperitoneal (i.p.) injection of tamoxifen at a dose of 80 mg/kg (1mg/mL) (Sigma) during a period of four days. To increase tumor burden in the colorectum triggered by *Lgr5*$^{eGFPCreERT2}$ driver, Cre recombination was induced by a single i.p. injection of tamoxifen (TAM; 8 mg/kg) followed by treatment with 1-3% dextran sodium sulfate (DSS) in drinking water for 5 consecutive days (Tanaka, 2012). Please note that *Lgr5*$^{eGFPCreERT2}$ transgene displays a mosaic expression in the intestinal epithelium with only about 10% of the normal crypts expressing GFP in the colon. Upon TAM and DSS treatment, >95% of GFP expressing cells in the colon were contained in adenomas and less than 5% remained in normal crypts (data not shown). Likewise, all adenomas arising in the colon of these mice were marked by the *LGR5*-GFP+ reporter. This observation implies that through this protocol we induced recombination of the *Apc*$^{fl/fl}$ in virtually all *Lgr5*$^{eGFPCreERT2}$ expressing cells. The general condition of animals were monitored using animal fitness, weight controls and tests to detect blood in stools (Haemoccult II®,SKD, France) throughout the experiment. When deteriorating clinical alterations were observed, the animals were culled. Experimental animals ranged from 2-4 month in age.

Detailed methods used for the rest of experimental procedures are included as supplementary material.

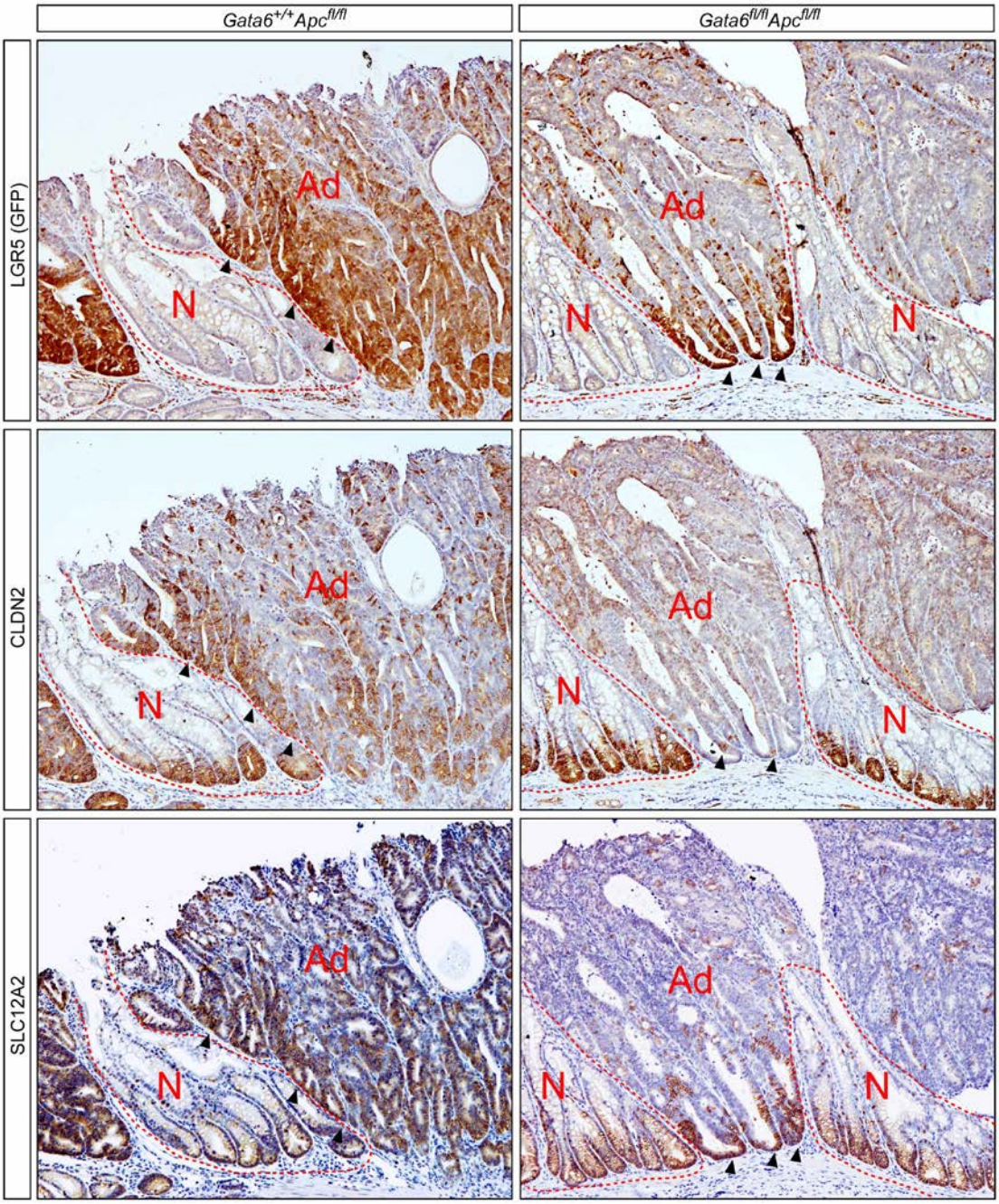
# SUPPLEMENTARY FIGURES



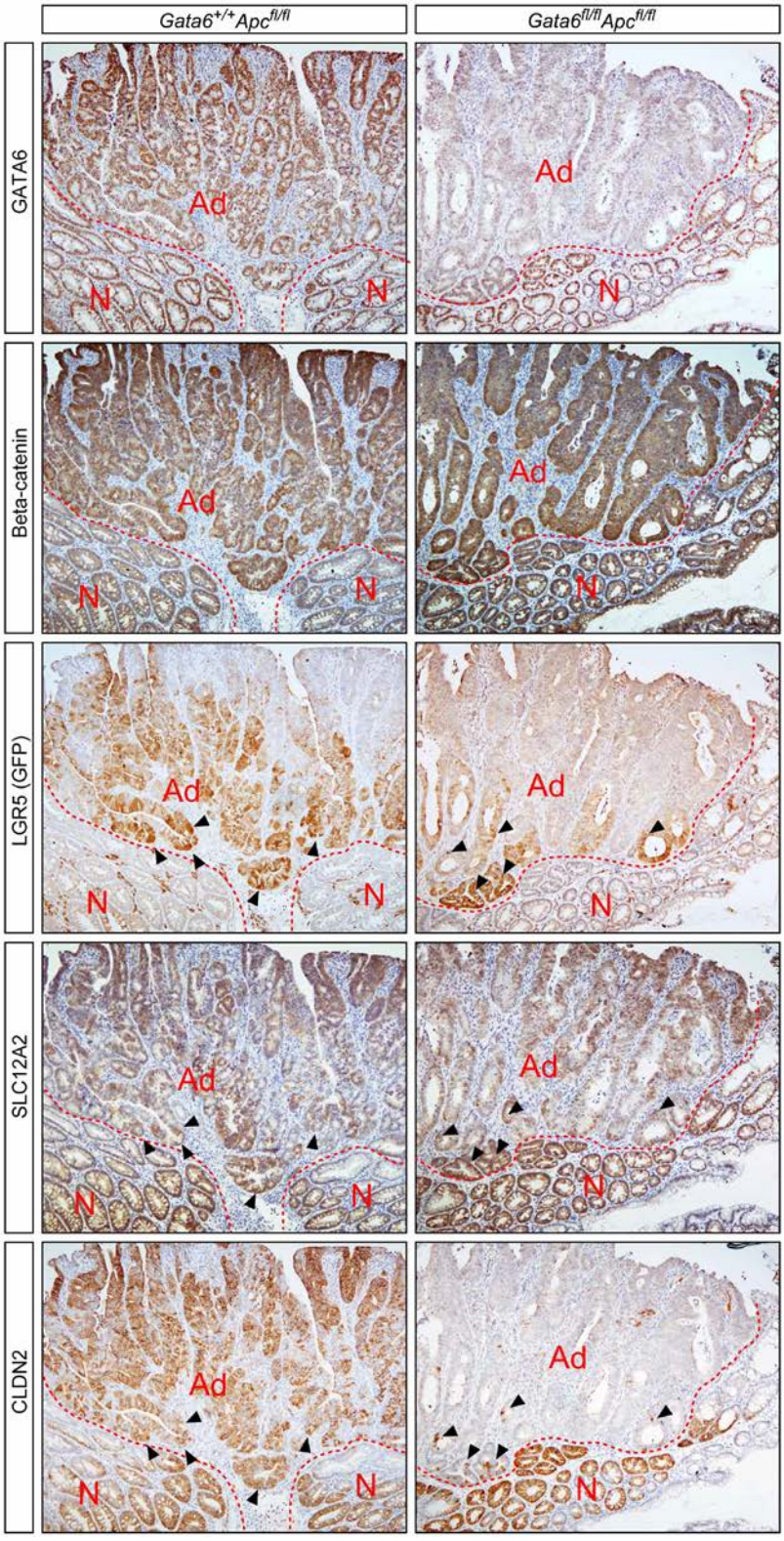

**Supplementary Figure S1. Additional ISC and AdSC genes identified in ARACNE network.** Representative stainings for CLDN2 (left) and SLC12A2 (right) in murine normal (upper) and tumor (middle and bottom) tissue. These markers have a restricted localization in the crypt base of normal and adenoma tissue.



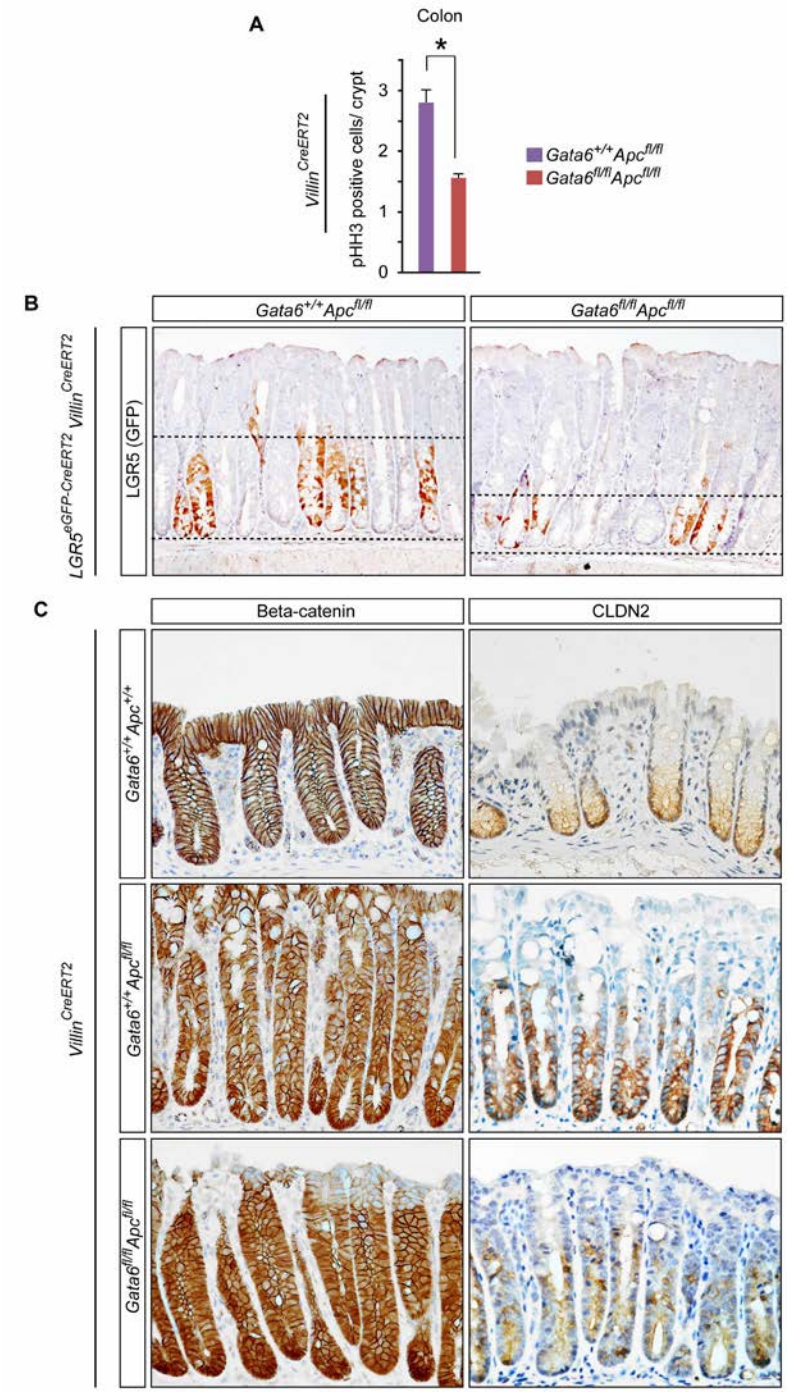

**Supplementary Figure S2. GATA6 is the only GATA factor expressed in colon.** (A) Representative stainings of GATA4 and GATA6 in adenomas from the adult mouse small intestine (top) and colon (middle) of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ mice as well as a human CRC patient sample (bottom). Note the lack of GATA4 staining in tumor tissue from the colon. Red dashed lines delimit normal (N) from adenoma (Ad) or CRC (T) tissues. (B) SW480 cells stably transduced with a doxycycline inducible vector to express *GATA4* or *GATA6*. Note introduction of either GATA factor resulted in upregulation of the ISC/AdSC markers *CLDN2* (orange) and *LGR5* (green). Fold-induction is the ratio between uninduced and doxycycline induced states for each GATA factor. Data are mean±s.e.m. of three independent experiments.



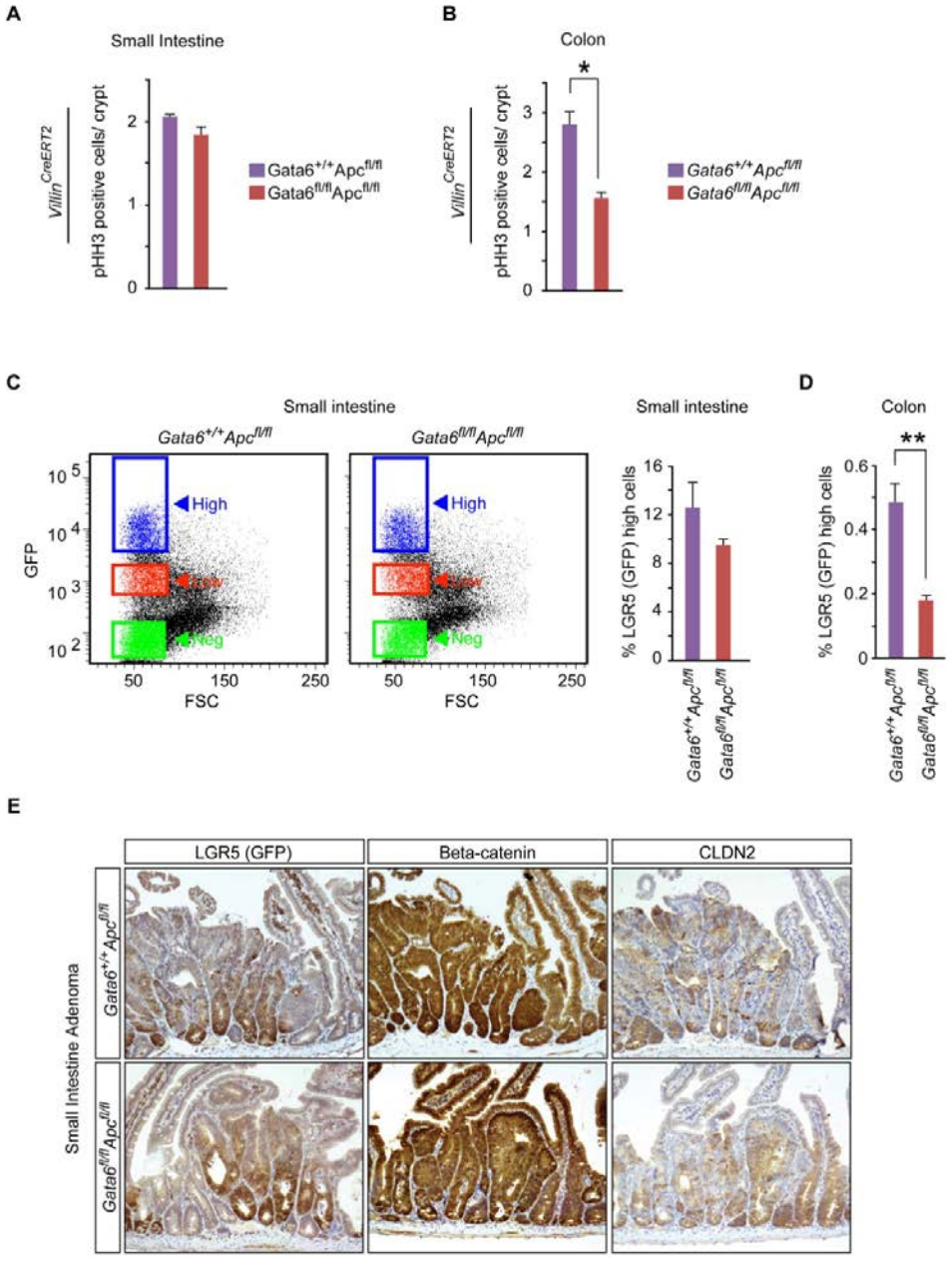

**Supplementary Figure S3. *Gata6* deletion in colon adenoma cells results in decreased CLDN2 and SLC12A2 expression.** Representative serial staining of LGR5 (GFP), CLDN2 and SLC12A2 in the same colonic adenomas shown in figure 2. Red dashed lines delineate normal (N) and adenoma (Ad) tissue. Cells expressing high levels of these three marker genes are positioned at the base of tumor glands (black arrowheads) in *GATA6* WT (left) adenomas. Notably, the number of cells expressing these markers is markedly reduced in compound mutants (right).



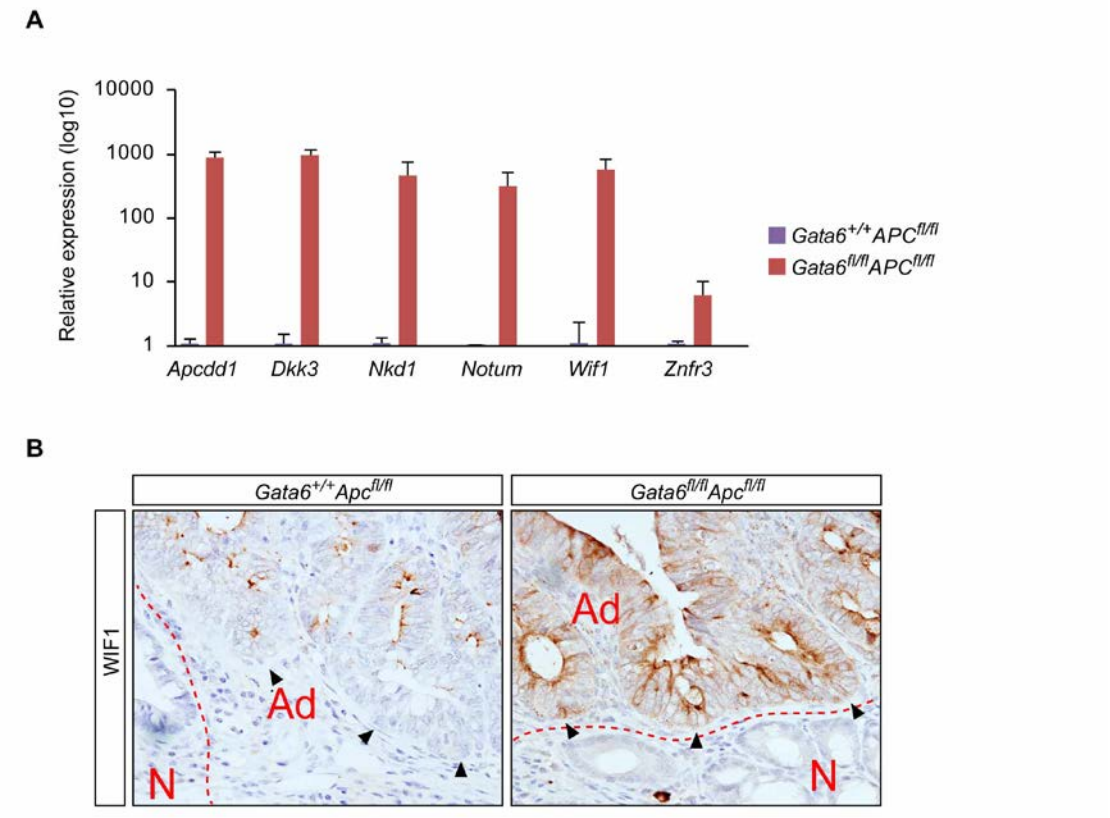

**Supplementary Figure S4. *Gata6* deletion in colon adenoma cells results in decreased AdSC markers (additional examples) .** Representative serial staining of markers GATA6, Beta-catenin, LGR5 (GFP), SLC12A2 and CLDN2 in colonic adenomas from *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right). Red dashed lines delineate normal (N) and adenoma (Ad) tissue. Arrowheads depict cells expressing high levels of these three AdSC marker genes at the bottom of the adenoma glands.



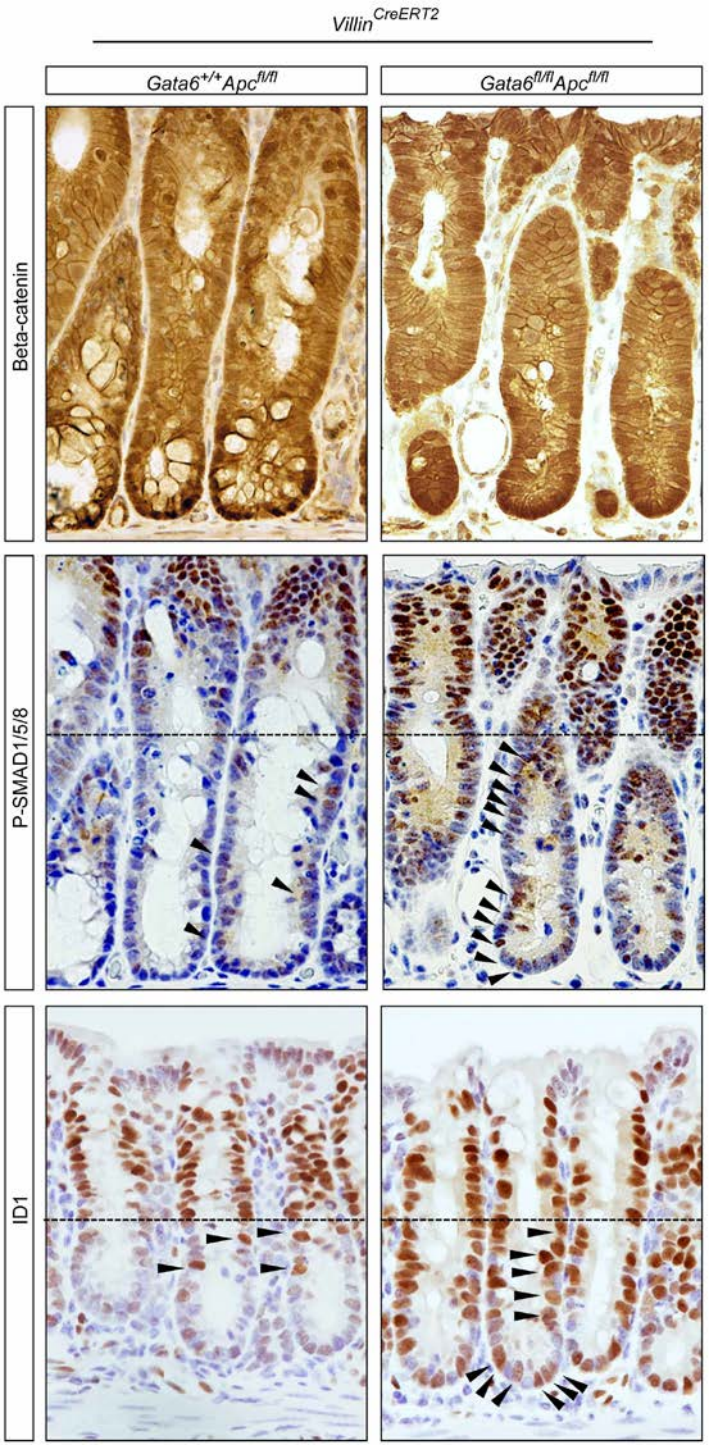

**Supplementary Figure S5. Conditional deletion of *Gata6* using the Villin CreERT2 driver results in proliferative defects and loss of AdSC markers LGR5 and CLDN2.** (A) Quantification of proliferating cells in the colon of *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bars) using phospho-histone-H3 (pHH3). Data are mean±s.e.m. of at least 100 crypts counted from 6 mice/genotype; Student t-test, *P<0.05. **(**B) Representative staining of LGR5 (GFP) from *Villin*$^{CreERT2}$*Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Villin*$^{CreERT2}$*Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) mice. Black dashed lines delineate AdSC zone. Note the mosaic expression of the Lgr5-GFP transgene in the colon. The size of the AdSC zone is reduced in compound mutants (dashed lines). (C) Representative staining of beta-catenin and CLDN2 in colonic crypts from *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{+/+}$ (top), *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (middle) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (bottom) mice 4 days after tamoxifen treatment (see methods). Note the sharp decrease in CLDN2 staining in compound mutants.



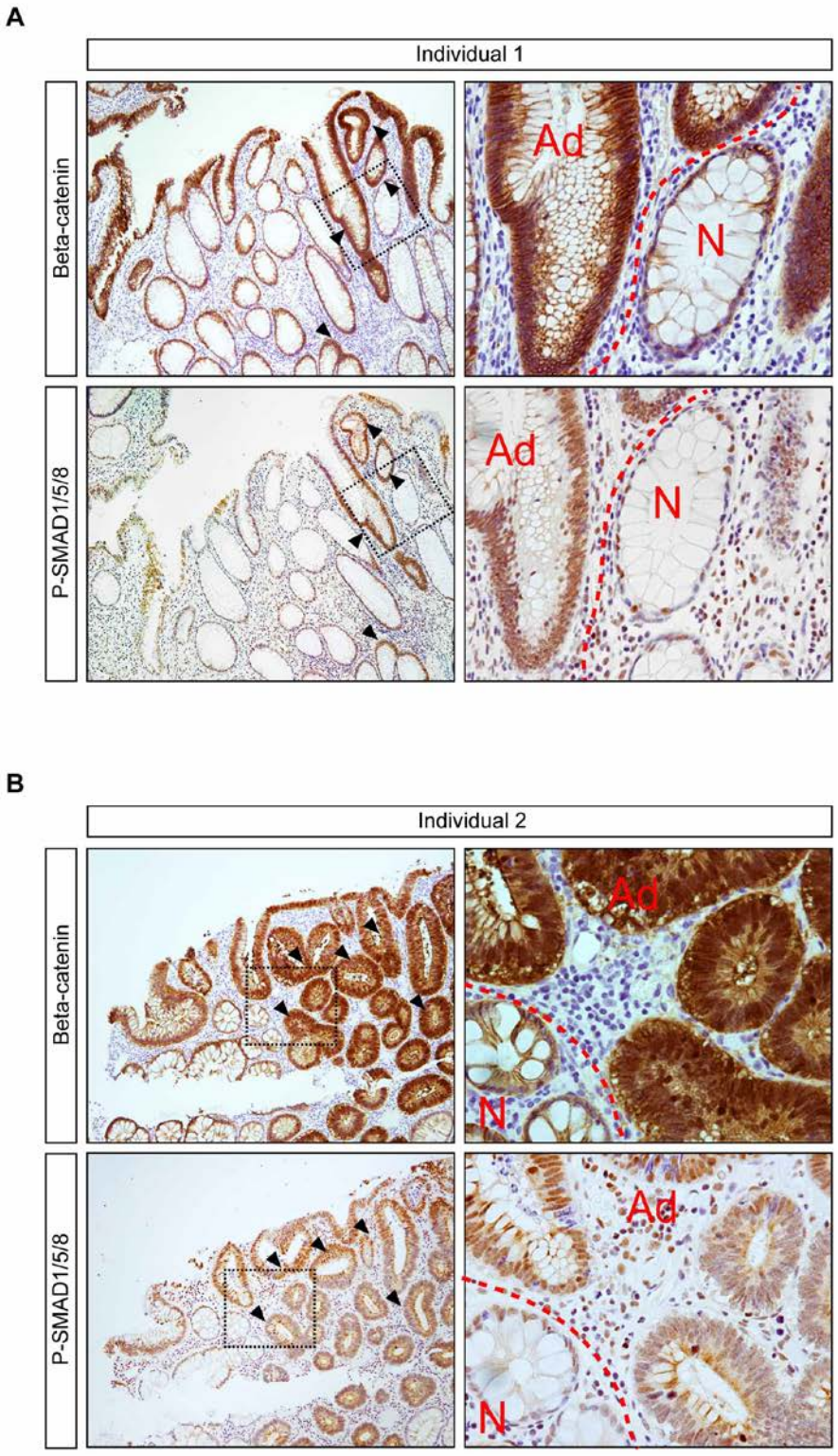

**Supplementary Figure S6. Effects of *Gata6* deletion in the small intestine compared to the colon. (**A-B) Quantification of proliferating cells, using phospho-histone-H3 labeling, in the small intestine (A) and colon (B) of *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bars) 4 days after Tamoxifen treatment (see methods). Note that only the colon displays a significant decrease in proliferating cells upon *Gata6* deletion. Data are mean±s.e.m. of at least 100 crypts counted from 6 mice/genotype; Student t-test, *P<0.05. (C) Epithelial cells from the small intestine of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice were disaggregated 4 days post tamoxifen administration and analyzed by flow cytometry. Single GFP high (red), low (green) and negative (blue) cells were gated using log GFP fluorescence intensity. Quantification of %GFP high cells from the small intestine and colon (D) of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bars) cells are shown. Data are shown as mean±s.e.m. of six mice; Student t-test, *P<0.05. (F) Representative stainings of LGR5 (GFP), beta-catenin and CLDN2 in adenomas from the small intestine of *Apc* null (top) and compound mutant (bottom) mice. Note that staining for all markers are comparable in the two genotypes.




**Supplementary Figure S7. Compound mutant tumor organoids highly upregulate WNT inhibitor genes. (**A) Differential expression of several WNT inhibitor genes between *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (purple bars) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (red bars) were validated by qRT-PCR. Data are shown as mean±s.e.m. of three independent experiments. **(**B) Representative staining of the WNT inhibitor WIF1 in *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) adenomas. Note increased WIF1 expression in compound mutants, at the base of adenoma glands (arrowheads).




**Supplementary Figure S8. Conditional deletion of *Gata6* using the Villin CreERT2 driver results in BMP pathway marker expansion in Adenoma crypts.** Representative staining of beta-catenin (top), P-SMAD1/5/8 (middle) and ID1 (bottom) in colonic crypts from *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ (left) and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ (right) mice. P-SMAD1/5/8 and ID1 are intensely expressed in adenoma cells from the upper half of crypts in *Apc* null (left panels). Note that compound mutants (right panels) exhibit expanded P-SMAD1/5/8 and ID1 domains that extend into the crypt base (arrowheads). Dashed lines identify the upper third of the crypt and arrowheads (black) point to P-SMAD1/5/8 and ID1 positive cells.




**Supplementary Figure S9. Adenomas from individuals with Familial Adenomatous Polyposis (FAP) stain strongly for marker of BMP activation P-SMAD1/5/8. (**A-B) Representative stainings of Beta-catenin (top) and P-SMAD1/5/8 (bottom) in human adenomas from FAP individual 1 (A) and 2 (B). Note that adenomatous tissue stains strongly with both markers. Red dashed lines delineate normal (N) and adenoma (Ad) tissue. Insets (black dashed boxes, left) of Beta-catenin (top, right) and P-SMAD1/5/8 (bottom, right).

# SUPPLEMENTARY TABLE LEGENDS

**Supplementary Table S1**. shRNA screen in LS174T colorectal cancer cells. Relative gene expression achieved with each individual shRNA knockdown for its targeted sequence (3$^{rd}$ column) and LGR5 (4$^{th}$ column) are shown. Target gene expression relative to control shRNA infected cells is shown. Each shRNA was performed in triplicate and representative values are shown. Screen was performed in LS174T cells infected with lentiviral vectors constitutively expressing the shRNA constructs outlines in the table. Relative expression of best two shRNAs against its target (highlighted in orange) and *LGR5* (highlighted in green). All other shRNAs from screen are shown in yellow. ND, not determined. Data available on included DVD.

**Supplementary Table S2**. *Gata6* tumor organoid microarrays. Differential gene expression between *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ tumor organoids after seven days of culture in control media. Fold changes (linear) and pvalues for all genes showing differential expression of at least 2-fold are shown. Normalized gene expression is derived four microarrays comprising to two independent biological replicates. Data available on included DVD.

**Supplementary Table S3**. GSEA output of *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ compared to *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ tumor organoids. Gene set enrichments are based on transcriptomics date from Table S2. Data available on included DVD.

# SUPPLEMENTARY METHODS

### ARACNe analysis to create the LGR5 Network

ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) is an algorithm which reverse engineers a gene regulatory network from microarray gene expression data. ARACNe uses mutual information (MI), an information theoretical measure, to compute the correlation between pairs of genes to infer a best-fit network of probable interactions (Basso et al., 2005; Margolin et al., 2006). In other word it calculates a score to assess whether two genes are related. To perform this analysis a minimum of 100 or so arrays are required (Basso et al., 2005; Margolin et al., 2006).
In the present study we have included 180 microarrays from human colorectal tumor tissues and cell lines. These arrays include 20 different commonly used colorectal cancer cell lines, 32 human adenomas (Ad) and 25 colorectal cancers (CRC) (Van der Flier et al., 2007). This algorithm is computationally demanding therefore a list of 2600 probes that represent putative WNT related genes were chosen to perform the analysis. These genes were analyzed against each other and the rest of probes from the microarray platform used to compute MI scores. The final ARACNe network analysis took two full weeks to using a Pentium 4 processor. An MI threshold of 0.105 with a pvalue of $1x10^{-7}$ and Data Processing Inequality (DPI) tolerance set to 0.1. A subset of this network was chosen for the shRNA screen and is represented in Table S1 and Figure 1A and B.

### Cell culture

DLD1, LS174T, SW48, SW480 and SW948 were purchased from the American Type Culture Collection (ATCC, USA). All cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with L-Glutamine (Invitrogen 41966, Paisley, Scotland) and 10% FBS (Invitrogen 10270) at 37ºC and 5% CO2.

### Plasmids, lentiviral vectors constructs and virus production

The Mission shRNA library (Sigma Aldrich) was used to perform shRNA screening of putative regulators of *LGR5* expression. A non-targeting shRNA vector (CV) (SHC002) was used as a control (Sigma Aldrich). A list of Mission reference numbers for the shRNAs used can be found in Supplemental Table 1. The efficiency of the knockdown was confirmed by qRT-PCR using TaqMan gene expression assays (Applied Biosystems).

Viral production was accomplished in 293T producer cells (ATCC) by transient transfection of the lentiviral construct with packaging plasmids pCAG-RTR2 (Rev-expressing), pCAG-VSVG (Envelope) and pCAG-KGP1R (Packaging) (Hanawa et al., 2002) and linear polyethylenimine (PEI) (Polysciences Inc. 23966, Warrington, USA) as the transfection reagent. Viral supernatants were collected at 48h and 72h post-transfection, filtered using 0.45µm PVDF filters, supplemented with polybrene (8µg/mL)

(Sigma H9268, St Louis, MO) and used to transduce target cells by O/N incubation. Infection efficiency of target cells was usually >90%, as determined by puromycin selection (2µg/mL, Invivogen, Toulousse, France). Cell lines stably expressing the corresponding constructs were maintained under antibiotic selection thereafter.
A total of 8 short-hairpin RNA sequences were designed against human *GATA6* gene. The short hairpins were cloned into doxycycline inducible pTRIPZ lentiviral vector (Open Biosystems). The most effective knockdowns were achieved with *GATA6* short hairpin 1 (shG6) and 2 (shG6_2). Oligonucleotide sequences used as template to create the vectors are shown below following a previously published protocol (Paddison et al., 2004).
*GATA6*_sh1:
(tgctgttgacagtgagcgacccagaccacttgctatgaaatagtgaagccacagatgtatttcatagcaagtggtctggg ctgcctactgcctcgga) and
*GATA6*_sh2:
(tgctgttgacagtgagcgcgcgctgacagaacgtgattcttagtgaagccacagatgtaagaatcacgttctgtcagcgc atgcctactgcctcgga), *GATA6*_sh3-8 are available upon request
Human *GATA4* (IMAGE: 8069086), *GATA6* and *GATA6* zinc finger mutant (Zhong et al., 2011), which lacks the eight most highly conserved bases of the zinc-finger motif, were subcloned from pcDNA3.1. HA tags were included in the N-terminus of each construct and all variants were first subcloned into pDONR221 and later cloned into pLENTI-PGK-HYGRO (Campeau et al., 2009) or were cloned in to the inducible pTRIPZ lentiviral vector (Open Biosystems) engineered to contain a Gateway cassette (Invitrogen) for doxycycline inducible expression of open reading frames (ORFs). Gateway (Invitrogen) cloning was achieved following manufacturer's instructions. GATA expression was induced in the presence of doxycycline (dox: 1 µg mL-1; Sigma Aldrich) for a period of 72h.

**RNA extraction, cDNA synthesis and quantitative real time-PCR**

Total RNA was extracted using the TRIzol® Plus RNA Purification Kit (Life Technologies). Briefly, CRC cell lines were scraped from cell culture dishes (Costar) and homogenized by pipetting in TRIzol solution (Life Technologies, 15596-018). After phase separation with chloroform, the upper aqueous phase was then mixed with 70% ethanol and bound to RNA columns (PureLink™ RNA Mini kit, Life Technologies) according to the protocol provided by the manufacturer. RNA was quantified using a Nanodrop spectrophotometer.
For qRT-PCR, the High Capacity cDNA Archive Kit (Applied Biosystems) was used to reverse transcribe 1 µg of purified RNA to cDNA according to the manufacturer's instructions. All qRT-PCR reactions were performed using Taqman probes (Applied Biosystems) and TaqMan Universal PCR Master Mix (Applied Biosystems). Human Taqman probes used in this study are the following: ABCC1 (Hs00219905_m1), ALDH1B1 (Hs00265114_s1), APCDD1 (Hs00537787_m1), ARL4C (Hs00255039_s1), ASCL2 (Hs00270888_s1), AXIN2 (Hs00610344_m1), BCAR3 (Hs00981962_m1), BCL11A (Hs00256254_m1), *BMP4* (Hs00370078_m1), C13orf18 (Hs00228336_m1), CADPS (Hs00186598_m1), CD46 (Hs00611256_m1), CDK6 (Hs01026372_m1), CDX1 (Hs00156451_m1), CDX2 (Hs00230919_m1), CHRM3 (Hs00327458_m1),

CLDN2 (Hs00252666_s1), CNN3 (Hs00156565_m1), Beta-catenin (Hs00994404_m1), CYP39A1 (Hs00213201_m1), CYP4X1 (Hs00380077_m1), DACH1 (Hs00362088_m1), DPEP1 (Hs01116752_m1), ENC1 (Hs00171580_m1), EPB41L2 (Hs00154988_m1), EPHA4 (Hs00177874_m1), EPHB2 (Hs00362096_m1), EPHB3 (Hs00177903_m1), ETS2 (Hs00232009_m1), FGFR4 (Hs01106908_m1), FOXQ1 (Hs00536425_s1), GATA4 (Hs01034629_m1), GATA6 (Hs00232018_m1), GGH (Hs00914163_m1), HES6 (Hs00610927_g1), HSD17B4 (Hs00264973_m1), HUNK (Hs0017191978_m1), ISYNA1 (Hs00375021_g1), KIAA1199 (Hs00378530_m1), KIAA1344 (Hs00394095_m1), KRT23 (Hs00210096_m1), LGR5 (Hs00173664_m1), MET (Hs00179845_m1), MSX2 (Hs00741177_m1), MYB (Hs00193527_m1), MYRIP (Hs00539278_m1), NFIX (Hs00231172_m1), NOX1 (Hs00246589_m1), NRXN3 (Hs00373450_m1), OXGR1 (Hs00369851_s1), PAH (Hs00609359_m1), PLEKHB1 (Hs00184204_m1), PPARGC1A (Hs01016719_m1), *PPIA (*Hs99999904_m1), PRDM16 (Hs_00922674_m1), PROX1 (Hs00160463_m1), PTPRO (Hs00243097_m1), RGMB (Hs00543557_m1), RGS19 (Hs00362370_m1), SATB1 (Hs00161515_m1), SATB2 (Hs00392652_m1), SLC12A2 (Hs00169032_m1), SLC35D3 (Hs01384745_m1), SLC39A10 (Hs00393794_m1), SLC6A6 (Hs00161778_m1), SLCO3A1 (Hs00203184_m1), SMOC2 (Hs00405777_m1), SP5 (Hs01370227_m1), TBX3 (Hs00195612_m1), TDGF1 (Hs02339499_g1) and TRAF5 (Hs01072224_m1).

Mouse Taqman probes used in this study are the following: *Apcdd1* (Mm01257559_m1), *B2m* (Mm437762_m1), *Bmp2* (Mm01340178_m1), *Bmp4* (Mm00432087_m1), *Bmp7* (Mm00432102_m1), *Dkk3* (Mm00443800_m1), Id1 (Mm00775963_g1), *Gata6* (Mm00802632_m1), *Id1* (Mm00775963_g1), Id3 (Mm00492575_m1), *Lgr5* (Mm00438890_m1), *Nkd1* (Mm00471902_m1), *Notum (*Mm01253273_m1), *Smad6* (Mm00484738_m1), *Wif1* (Mm00442355) and *Znrf3 (*Mm01191453_m1). StepOnePlus Real-Time PCR and ABI Prism 7900 Sequence Detector Systems were used to carry out the qRT-PCR reactions in clear optical 96 and 384-well reaction plates with optical covers, according to manufacturer's instructions. Gene expression levels were normalized using the endogenous control *PPIA (*Human) and *B2m* (Mouse) for each sample and differences in target gene expression were determined using SDS 2.4 or StepOne 2.2 plus software. Error bars represent standard error of the mean (s.e.m) of samples performed in triplicate.

**Microarrays of adenoma tumor organoids**

Total RNA for biological replicates of *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ adenoma tumor organoids grown for one week in control media (see section In vitro culture of normal colon and colon adenoma tumor organoids below for details). RNA was quantified using a Nanodrop spectrophotometer, quality was assessed using a Bioanalyzer (Agilent) and high quality samples (RIN ≥9) were used as template to hybridize to Affymetrix Mouse Genome 430 PM Strip microarrays (Affymetrix) in our core facility (IRB Transcriptomics) using standard techniques.

## Tissue processing

Main text outlines genotypes and treatment regiments used during experimentation with the varying mouse models used in this work.
Whole intestines were removed cut longitudinally and rinsed well with PBS. The small intestinal was cut into 3 equal parts and the colon was kept whole. Murine tissue was fixed on Whatman paper O/N at RT in 10% neutral buffered formalin solution, bundled in plastic cassettes and embedded into paraffin blocks.

## Tamoxifen stock preparation

Briefly, tamoxifen was rapidly (<2min.) dissolved in 100% ethanol (1/10$^{th}$ of the final volume) at 55C. Once completely dissolved, solution was immediately combined with 9/10$^{th}$ of corn oil. The emulsion was mixed by inversion to achieve a translucent solution. Aliquots were stored at -20C.

## Colon and small intestinal adenomatous tissue purification

For flow cytometry experiments using the *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice, four daily injections of (80mg/kg) tamoxifen were administered before cell isolation. All other experiments performed with Lgr5$^{eGFPCreERT2}$ involved a single injection of tamoxifen (8mg/kg) followed by DSS treatment to enhance tumor burden in the colon. Tissue was disaggregated, as outlined below, and GFP+ cells were analyzed from adenomatous tissue using GFP fluorescence. LGR5$^{high}$ tumor organoid cultures were isolated and grown using the same approach outline above. Conversely, the *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice were sacrifice without tamoxifen treatment and recombination was effectively induced *in vitro* using 100ng/mL 4OHT during a period of 3 days. Tumor organoids grown from either *Lgr5*$^{eGFPCreERT2}$ or *Villin*$^{CreERT2}$ animals were indistinguishable from one another and *Gata6* null tumor spheres in both mouse backgrounds gave rise to crypt-like folds in the absence of BMP pathway inhibitors. Notably, *Lgr5*$^{eGFPCreERT2}$ animals required tamoxifen induction *in vivo*, as the Cre in this model functions poorly *in vitro*.
Briefly, Mouse small intestine and colon were freshly isolated and cut longitudinally to expose the epithelial sheet. Segments of both the colon and small intestine were washed in ice cold HBSS and subsequently incubated in 8mM EDTA in HBSS for 15 min in a water bath at 37C. Colons were collected 0.5cm away from the anus. Samples were then carefully washed in ice cold PBS, to remove EDTA, then vigorously shaken for 60s to obtain a supernatant enriched for crypts. The supernatant was then decanted into a new tube and placed on ice. The procedure was repeated a second time to increase crypt enrichment and homogenize quantity of disaggregated crypts between samples. Crypt supernatants were then spun at low speed (500rpm) for 5 min, to remove contaminating single cells and debris from crypts fractions. Supernatants were decanted from the pelleted crypts and crypts were washed twice with ice cold PBS. The final wash in PBS was increased to 1200rpm to concentrate the pelleted crypts. Crypt suspensions were kept on ice until seeded in cell culture in the case of *Villin*$^{CreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Villin*$^{CreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice.

In the case of *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{+/+}$*Apc*$^{fl/fl}$ and *Lgr5*$^{eGFPCreERT2}$*Gata6*$^{fl/fl}$*Apc*$^{fl/fl}$ mice, adenomatous tissue was enzymatically disaggregated (0.4mg/mL Dispase, in HBSS) for 30 min at 37ºC using an orbital shaker. After disaggregation, 5% FBS was added to neutralize the Dispase (Invitrogen 17105) and cell aggregates were mixed by syringing (23 gauge) the cell suspension 5 times. Single cell suspensions were passed through a 40 µm mesh filters, remove any unwanted debris that might clog the FACS. Cells were then spun (1200 rpm for 5 min at 4ºC) and resuspended in HBSS containing 5% FBS. Fluorescence in the Blue channel (GFP) was gated to clearly separate GFP$^{high}$ and GFP$^{low}$ populations. A total of six colons and small intestines were used to quantify differences in GFP$^{high}$ cell populations.

### *In vitro* culture of normal colon and colon tumor organoids

Freshly isolated crypts were counted and mixed at a density of 300 crypts units per 30µLof Matrigel GFG (BD Biosciences 356231) on ice. Matrigel droplets of 30µLwere carefully dispensed in the centre of pre-warmed standard 48 well cell culture plates (Costar 3548). After solidification (10 min at 37°C), crypts were overlaid with 250µL CoSC media: 50% advanced DMEM (Life Technologies, 12491-023), 50% Wnt3a-conditioned medium from producer L-cells (or Wnt3a Millipore 10ug/mL), 1X Glutamax (Gibco 35050), 1X HEPES (Gibco 1M-15630), 1X N-2 (Life Technologies 17502-048), 1X B-27 (12587-010), 1 mM N-Acetyl-L-cysteine (NAc) (Sigma Aldrich A9165), 50ng/mL EGF (Life Technologies PHG0311), 1ug/mL RSPO1 (Jung et al., 2011) (produced in house, see REF 8 for details), 100ng/mL noggin (Peprotech 120-10C), 1ug/mL Gastrin (Sigma Aldrich G9020) and 10µMRock inhibitor Y-27623 (Merk 688000). Note: Y-27623 is only required when crypts or single cells are freshly seeded. Once tumor organoids have started to grow, typically by day 3, media is supplemented with 10nM/mL 4-hydroxytamoxifen (4OHT) (Sigma H7904) to induce Cre mediated recombination. When 4OHT colon tumor organoids were passaged, they were grown in Adenoma media (Control media): 50% advanced DMEM (Life Technologies, 12491-023), 1X Glutamax (Gibco 35050), 1X HEPES (1M Gibco 15630), 1X N-2 (Life Technologies 17502-048), 1X B-27 (12587-010), 50ng/mL EGF (Life Technologies PHG0311) and 10µMRock inhibitor Y-27623 (Merk) when freshly seeded, as with CoSC media. Media replacement was performed every 2 days. Cultures were typically passaged every 7-10 days. When passaging either normal colon or adenoma tumor organoids, we gently washed the Matrigel drops with PBS. Each drop was then resuspended in PBS by mechanical disruption of the drop to free tumor organoids from the Matrigel matrix. Centrifugation was performed at low speed (500 rpm) for 3 minutes to separate tumor organoids from single cells and debris. Pelleted tumor organoids were then mixed with 0.5X trypsin (Life Technologies) and diluted with PBS for approximately 5 minutes, at 37°C. Once cells were disaggregated trypsin was neutralized by addition of 1 volume of advanced DMEM supplemented with HEPES and Glutamax. Centrifugation was performed at 1200 rpm for 5 minutes to pellet single cells. All centrifugation steps and handling of tumor organoids after disaggregation were performed at 4°C. *Apc* null tumor organoids grown in control media have been passaged for > 8 month in the absence of BMP inhibitors with no obvious deterioration in organoid formation efficiency.

**Organoid growth kinetics and serial passages**

Organoid cultures were disaggregated to single cells and counted with a Neubauer chamber. Serial dilutions (from single cells suspensions) were embedded into 25µLdrops of Matrigel (BD Biosciences). Using the CellTiter-Glo luminescent cell viability assay kit (Promega PR-G7572 ) growth kinetics of organoid cultures were determined according to the manufacturers recommendations, with the exception that Matrigel drops had to be carefully disrupted by pipetting (10x/well) in the prepared CellTiter-Glo solution before luminescence was determined using a Luminometer.

**Survival analysis**

Kaplan-Meier survival curves were produced using the log rank test. Statistical analyses were performed using GraphPad Prism (GraphPad Software Inc.).

**Colon tumor burden analysis using ImageJ and MosaicJ**

Cre recombination was induced in the experimental animals by a single i.p. injection of tamoxifen (8 mg/kg) followed by treatment of 1-3% dextran sodium sulfate (DSS) in drinking water for 5 days. The general condition of animals were monitored using animal fitness, weight controls and tests to detect blood in stools (Haemoccult II®,SKD, France) throughout the experiment. When deteriorating clinical alterations were observed, the animals were culled. After O/N fixation in 10% neutral buffered formalin solution and washed with 1X PBS, sequential pictures of colon tissues were taken using a MVX10 macroscope (Olympus). Image manipulation and analyses were performed using ImageJ (http://rsbweb.nih.gov/ij/). Using MosaicJ, multiple high resolution images of the colon were combined to create a mosaic of each colon sample. Once assembled, the whole colon surface area was calculated using ImageJ features. DSS treatment mainly exerts effects in the distal half of the colon, thus tumor areas were calculated as a percentage of colon area.

**Immunohistochemistry (IHC)**

Whole murine small intestine and colon were freshly fixed in 10% neutral buffered formalin (Sigma Aldrich) O/N and embedded in paraffin blocks.

Adenoma tumor organoids were washes with PBS and fixed in the Matrigel drop for 4h with 10% neutral buffered formalin. Each drop was then resuspended in PBS by mechanical disruption of the drop to free tumor organoids from the Matrigel. Centrifugation was performed at 1200 rpm for 5 minutes to pellet tumor organoids For each experimental condition the lower third of 1.5mL conical tubes was cut using scissors and placed upside-down on the bench. Tumor organoids were mixed with hot agarose to achieve a final agarose concentration of 1%. The agarose/tumor organoid slurry was immediately added to the inverted 1.5mL conical tubes to form conical agarose plugs. Agarose plugs were then fixed 10% neutral buffered formalin O/N and embedded in paraffin blocks.

Immunostainings were performed using 4µm tissue sections according to standard procedures. Briefly, antigen retrieval was performed, samples were blocked with Envision FLEX Antibody Diluent (Dako, K8006) for 1h at RT, and then the primary antibodies were incubated in blocking solution (see table below for details). Slides were washed with PBS (or TBS for phospho antibodies) and the corresponding secondary antibody was incubated with the sample for 45 min at RT. Samples were developed using DAB, counterstained with haematoxylin and mounted. Primary antibody details, dilutions, retrieval, blocking and incubation conditions are outlined in the table below.

**Immunofluorescence (IF)**

Tissues examined by fluorescence Immunostaining (IF) were prepared and processed as above and outlined in the table below, although to perform double and triple immunostainings with Id1some antibodies from the table below were substituted. The primary antibodies used for IF were Goat anti-GFP (Abcam, ab5450) (1:100), Mouse anti-beta-catenin (Sigma, C7207) (1:100) and Rabbit anti-Id1 (Biocheck, BCH-1/37-2) (1:100). Briefly, Immunostaining on paraffin embedded sections were performed using 4µm tissue sections according to standard procedure outlined above for standard IHC. Samples were then blocked with Envision FLEX Antibody Diluent (Dako, K8006) for 30 min and incubated with the primary antibodies overnight at 4°C followed by incubation with the secondary antibodies for 1h at RT. Nuclei were stained with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) (Sigma, D9542) at 0.5µg/mL for 15min. Slides were mounted in DAKO Fluorescent Mounting Medium (DAKO, S3023) or Vectashield (Vector labs H-1000 Burlingame, CA). The secondary antibodies used were Donkey anti-Rabbit 649 (Dylight, -20ºC), Donkey anti-Mouse 594 (Dylight, -20ºC) and Donkey anti-Goat 488 (Dylight, -20ºC) (Jackson Immunoresearch). All primary antibodies were diluted 1:100 and secondary antibodies were diluted 1:400 with Envision FLEX Antibody Diluent.

| **Primary antibody** | **Manufacturer and Reference** | **Dilution** | **Retrieval, blocking, primary antibody incubation** |
|---|---|---|---|
| Goat anti-GATA6 | R&D (AF1700) | 1:500 | Citrate pH6, autoclave 20 min, 5% BSA 0.1% Tween, O/N at 4ºC |
| Rabbit anti-GFP | Abcam (ab6556-25) | 1:1000 | Citrate pH6, autoclave 20 min, Envision FLEX Antibody Diluent, O/N at 4ºC |
| Mouse anti-Beta-catenin | SIGMA (7207) | 1:100 | Citrate pH6, autoclave 20 min, 0.05% BSA, 20 min at RT |
| Rabbit anti-P-SMAD1/5/8 | Cell Signaling (9511S) | 1:500 | Citrate pH6, boil 20 min, 5% Envision FLEX Antibody Diluent, O/N at 4ºC |
| Goat anti-GATA4 | Santa Cruz (sc-1237) | 1:1000 | Citrate pH6, autoclave 20 min, 1% BSA, O/N at 4ºC |
| Rabbit anti- | Upstate | 1:3000 | Citrate pH6, autoclave 20 min, 5% |

| | | | |
|---|---|---|---|
| P-Histone H3 | (06-570) | | milk, O/N at 4ºC |
| Rabbit anti-CLDN2 | Abcam ab15100 | 1:100 | Citrate pH6, autoclave 20 min, Envision FLEX Antibody Diluent, O/N at RT |
| Rabbit anti-SLC12A2 | Sigma HPA020130 | 1:2000 | Citrate pH6, autoclave 20 min, Envision FLEX Antibody Diluent, O/N at 4ºC |
| Rabbit anti-ID1 | BioCheck BCH-1/195-14 | 1:4000 | Tris-EDTA pH9, boil 20 min, Envision FLEX Antibody Diluent, 2h at RT |
| Rabbit anti-WIF1 | R&D (AF135) | 1:100 | Citrate pH6, autoclave 20 min, Envision FLEX Antibody Diluent, O/N at 4ºC |

**ChIP and ChIP-seq**

Cells were cross-linked with 1% formaldehyde for 15 min at room temperature, harvested in lysis buffer ($2 \times 10^7$ cells/ml) and sonicated on Covaris system (shearing time 30 min, 20% duty cycle, intensity 10, 200 cycles per burst and 30 s per cycle) in a volume of 2 ml. ChIP was then performed with standard protocol, using GATA6 (R&D AF1700), beta-catenin (BD 610154), TCF4 (Santa Cruz sc-8631) and p300 (Santa Cruz sc-585) antibodies. For ChIP-seq, 20ng of DNA, as quantified by fluorometry, was resolved by electrophoresis and fractions of 50-250bp were extracted. Fractions were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters as in Illumina's "TruSeq DNA Sample Preparation Guide" (part # 15005180 Rev. C). Adapter-ligated libraries were completed by limited-cycle PCR with Illumina PE primers (12 cycles). The resulting purified DNA library was applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v5) and sequenced on the Genome Analyzer IIx with SBS TruSeq v5 reagents by following manufacturer's protocols. Image analysis and per-cycle base calling was performed with Illumina Real Time Analysis software (RTA1.13). Conversion to FASTQ read format was performed with CASAVA-1.8 (Illumina).

Sequence alignment to the reference genome (GRCh37/hg19, February 2009) was made with BWA version 0.5.9-r16 algorithm allowing up to 0/1 mismatches, using default settings. Only non-duplicate reads were kept for the analysis and ChIP-seq and Input alignments files were normalized to the same number of reads randomly before the peak detection. To visualize the data in the University of California Santa Cruz genome browser (http://genome.ucsc.edu) reads were directionally extended to 300 bp, and for each base pair in the genome the number of overlapping sequence reads was determined and averaged over a 10 bp window to create a wig file.

Peak finding and motif discovery. Significant peaks of GATA6 binding were identified with macs27 v2.0.9 using q-value<0.05 and the others settings by default.

Reproducibility of replicates was assessed ranking the peaks of each by q-value. For each replicate, top 40% of peaks were evaluated to be present in the list of identified peaks for the other replicates using a previously described method (Chikina and Troyanskaya, 2012). Such as for all replicates more of the 85% of the top 40% significant peaks were present in the others and the number of peaks identified for each replicate don't differ by more than a factor of 2, reads from replicates were combined to repeat the peak finding analysis as described above, significant peaks obtained following this strategy were used for further analysis.

PeakAnalyzer software was assessed to identify functional elements proximal to the GATA6 peaks. Motifs were identified with the MEME suite, using significant peaks, and then TOMTOM was used to compare the identified motifs with known transcriptional motifs.

Campbell K, Whissell G, Franch-Marro X, Batlle E, Casanova J. Specific GATA factors act as conserved inducers of an endodermal-EMT. Dev Cell. 2011; 21(6):1051-61. DOI: 10.1016/j.devcel.2011.10.005

Follow link below

http://www.cell.com/developmental-cell/retrieve/pii/S1534580711004229

# CHAPTER 3

# SPECIFIC GATA FACTORS ACT AS CONSERVED INDUCERS OF AN ENDODERMAL-EMT

Kyra Campbell[1,2], Gavin Whissell[2], Xavier Franch-Marro[3], Eduard Batlle[2,4] and Jordi Casanova[1,2]*.

1. Institut de Biologia Molecular de Barcelona-CSIC, Parc Cientific de Barcelona, 08028 Barcelona, Spain
2. Institut de Recerca Biomèdica de Barcelona, Parc Cientific de Barcelona, 08028 Barcelona, Spain.
3. Institut de Biologia Evolutiva CSIC-UPF, Passeig de la Barceloneta, 37-49, 08003 Barcelona, Spain.
4. Institució Catalana de Recerca i Estudis Avançats (ICREA).

Correspondence should be addressed to:
Jordi Casanova (jcrbmc@ibmb.csic.es)

## SUMMARY

The epithelial-to-mesenchymal transition (EMT) converts cells from static epithelial to migratory mesenchymal states (Hay, 1995). Here, we demonstrate that EMT in the *Drosophila* endoderm is dependent on the GATA-factor Serpent (Srp), and that Srp acts a potent trigger for this transition when activated ectopically. We show that Srp affects endodermal-EMT through a downregulation of junctional dE-Cadherin (dE-Cad) protein, without a block in its transcription. Moreover, the relocalization of dE-Cad is achieved through the direct repression of *crumbs (crb)* by Srp. Finally, we show that hGATA-6, an ortholog of Srp, induces a similar transition in mammalian cells. Similar to Srp, hGATA-6 acts through the downregulation of junctional E-Cad, without blocking its transcription, and induces the repression of a Crumbs orthologue, *crb2*. Together these results identify a set of GATA factors as a conserved alternative trigger to repress epithelial characteristics and confer migratory capabilities on epithelial cells in development and pathogenesis.


## INTRODUCTION

The epithelial-to-mesenchymal transition (EMT) plays crucial roles during development, and when activated inappropriately can promote tumor progression and metastasis. This process converts cells from static epithelial to migratory mesenchymal states (Hay, 1995); (Thiery et al., 2009) and is often carried out to different extents, as many cells adopt mesenchymal behaviours yet retain some epithelial features (Revenu and Gilmour, 2009). One of the best-known examples of an EMT underlies the formation of the *Drosophila* mesoderm, which is triggered by Snail (Sna). Subsequent to its discovery in *Drosophila*, genes of the *sna* family were found to induce EMT during the development of many organisms and inappropriately activated in several forms of cancer. A similar transition occurs during the formation of the *Drosophila* endoderm, when epithelial cells convert to a migratory mass of mesenchymal cells (Campos-Ortega and Hartenstein, 1985; Reuter, 1994; Skaer, 1993; Tepass and Hartenstein, 1994b), which migrate through the embryo and later re-epithelialize to give rise to a large portion of the intestinal tract. However, in contrast to the mesoderm, this transition seems to occur independently of Sna (Reuter and Leptin, 1994; Tepass and Hartenstein, 1994b).

Formation of the *Drosophila* intestinal tract originates from two groups of cells at each pole of the blastoderm. In particular, cells from the posterior region

give rise to the hindgut, of ectodermal origin, and the posterior midgut (PMG), of endodermal origin (Campos-Ortega and Hartenstein, 1997). While derived from a common primordium, hindgut and PMG cells undergo dramatically different cell behaviours during development. Hindgut cells remain epithelial and relatively static throughout, whereas PMG cells temporarily adopt mesenchymal behaviour and initiate migration towards the center of the embryo, where they fuse with the anterior midgut (AMG) to form a continuous intestinal tract (Campos-Ortega and Hartenstein, 1997; Reuter et al., 1993) Supplementary Movie S1). The decision of whether to form hindgut or the PMG is regulated by the activity of a single gene product, the GATA factor Serpent (Srp), which is active in the cells giving rise to the PMG. In *srp* mutants a homeotic transformation has been reported as an ectopic hindgut forms at the expense of the PMG (Reuter, 1994).

Here we investigate the role of Srp during endoderm development and show that Srp is both necessary and sufficient to induce epithelial cells to undergo a transition to a non-apicobasally polarised, motile state. At the molecular level, this transition involves the repression of genes encoding the apically localised proteins Crumbs (Crb), Stardust (Sdt) and Stranded-at-second (SAS). In contrast, while adherens junction associated proteins are delocalised, they are not transcriptionally repressed. We show that rather than abolishing *dE-Cadherin (dE-Cad)* gene transcription, Srp acts at the level of dE-Cad protein and regulates its membrane localisation. Moreover, this relocalization of dE-Cad requires the direct repression of *crumbs (crb)* by Srp. We extend these results by showing that human GATA4 and GATA6 can also induce an epithelial to non-polarised migratory cell transition, and similarly this occurs in an E-Cadherin (E-Cad) transcription independent manner. Furthermore, overexpression of GATA6 induces the transcriptional repression of a Crumbs orthologue, *crb2*. These results unveil the role of specific GATA factors as inducers of an EMT that is characterized by affecting E-Cad protein localisation rather than transcription.

# RESULTS

### PMG cells undergo an epithelial transition characterised by a loss of apico-basal polarity and downregulation of junctional dE-Cad

To better understand the endoderm transition, we analysed markers for cell polarity, adhesion and the cytoskeleton prior to, and during PMG migration. Analysis of F-actin (F-Act) and apico-basal markers reveals that in stage 9 embryos, endoderm cells are fully polarised epithelial cells (Figure 1A,D,G,J,M).

Electron microscopy studies showed that posterior endoderm cells possess complete zonula adherens during stages 6-9 (Tepass and Hartenstein, 1994a); accordingly we find a tight apical localisation of dE-Cad in endoderm cells throughout these stages (Figure 1P and not shown). During stage 10, a loss of epithelial characteristics is first apparent in cells at the most distal region of the PMG, and later throughout the PMG. Cells lose their columnar appearance, become more rounded and irregular in shape (Figure 1B,C,N,O). Apical and junctional proteins are gradually lost from the cell surface, and the overall levels of these proteins dramatically decrease (Figure 1F,I,L,R). When PMG cells initiate migration, apical and sub-apical proteins Stranded-at-second (SAS),

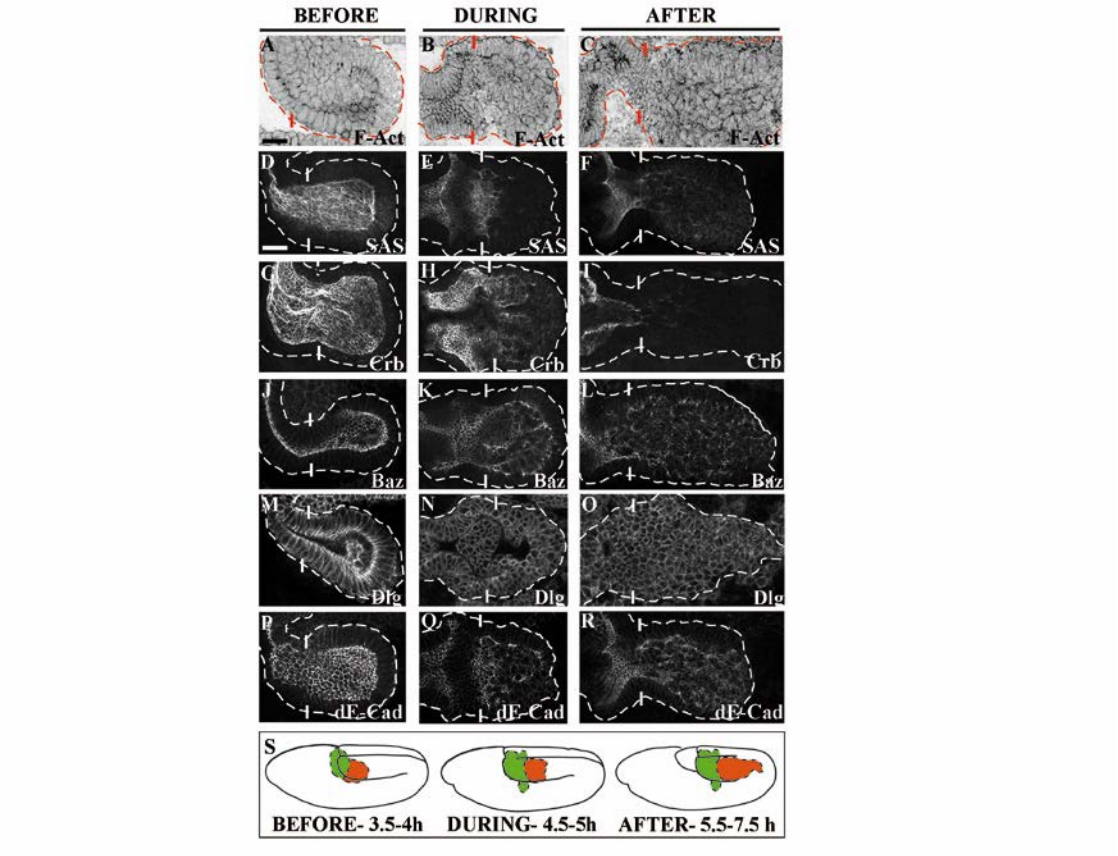



**Figure 1. As endodermal cells undergo EMT they lose epithelial morphology, apico-basal polarity and downregulate junctional dE-Cad.** A-R, Wild-type embryos stained for the cytoskeletal marker F-act (A-C), the apical proteins SAS (D-F), Crb (G-I) or Baz (J-L), the basolateral marker Dlg (M-O) or the adherens junction marker dE-Cad (P-R). Before is 3.5-4 hours after fertilization (stage 9), During, 4.5-5 hours (stage 10) and After 5.5-7.5 hours (stage 11-early stage 12). The PMG is outlined by dotted lines, vertical lines demark the separation between the hindgut and PMG during stage 9, first column, and between the Malpighian tubules and PMG during stages 10-11, second and third columns. The PMG is always to the right of the vertical lines. Scale bar, 20 μm. (S) Schematic showing the position of the PMG (red) with respect to the hindgut/Malpighian tubule primordium (green) and the rest of the embryo during stages 9-12. See also Supplementary Figure S1; page 100 and Supplementary Movies S1 and S2.

Crumbs (Crb) and Stardust (Sdt) are no longer detectable, while low levels of junctional dE-Cad, Bazooka (Baz) and atypical protein kinase C (aPKC) remain (Figure 1F,I,L,R and Supplementary Figure S1; page 100). Indeed dE-Cad, Baz and aPKC lose their tight apical localisation and instead co-localise to dynamic punctate accumulations at the cell membrane (Supplementary Figure S1; page 100, Supplementary Movie S2).

Electron microscopy studies previously showed that after adherens junctions fragment in the endoderm, low levels of spot adhesion junctions are present throughout the PMG (Tepass and Hartenstein, 1994a). We suggest that dE-Cad/Baz/aPKC could localise to spot adhesion junctions, which may act as dynamic adhesions facilitating cohesive migration of the mesenchymal cell mass.

**Srp activity is required for endodermal-EMT in the PMG**

Formation of the PMG depends on *srp* (Reuter, 1994) and occurs independently of *sna* genes (Reuter and Leptin, 1994; Tepass and Hartenstein, 1994b) Supplementary Figure S2; page 101), conversely to AMG formation that requires both *srp* and *sna* (Reuter and Leptin, 1994). Thus, to analyze the role of Srp in endoderm transition we have focussed our analysis on the PMG. In *srp* mutants endodermal cells retain epithelial shape and polarity (Figure 2A,B) and dE-Cad remains tightly localised to the apical domain (Figure 2B). Thus the endodermal-EMT depends on Srp activity.

**Maintenance of Srp activity prevents PMG cells from reforming an epithelium**

Next, we looked to see when Srp protein is expressed in the PMG, and found that it is first expressed during stage 5 (data not shown), indicating that there is a time-lag of 2 hours before endodermal EMT becomes visible at the morphological level. This suggests that a certain threshold of Srp may need to be reached for it to become active, or that its activity is being repressed by an unknown mechanism during stages 5-9. At the stage when we see a downregulation of Crb and Sdt, and delocalisation of junctional proteins, Srp is expressed very strongly in the PMG (Figure 2c arrows), however during stages 11-12 there is a gradual loss of Srp expression in the PMG (Figure 2D,E arrows). As Srp starts to disappear from the cells as they start migrating we reasoned that the temporally restricted expression of Srp could be related to the fact that PMG cells only undergo a transient transition to mesenchymal behaviour; once endodermal cells have migrated to their final position, they

undergo a reverse transition, mesenchymal-to-epithelial, forming the midgut epithelium (Figure 2F,H,J,L and (Tepass and Hartenstein, 1994b). To test this we provided sustained Srp expression at the time when it would normally be turned off (see Experimental Procedures). Under these circumstances, the midgut epithelium does not properly form and we detect multi-cell layered stretches instead of a normal mono-layered epithelial morphology (Figure 2G,I). Analysis of several markers reveal that continued expression of Srp prevents PMG cells from re-establishing epithelial characteristics (Figure 2H-M). Notably, under these conditions dE-Cad was almost undetectable in PMG cells, (Figure 2M). Altogether, these results indicate *srp* has to be turned off for endodermal cells to regain epithelial characteristics.

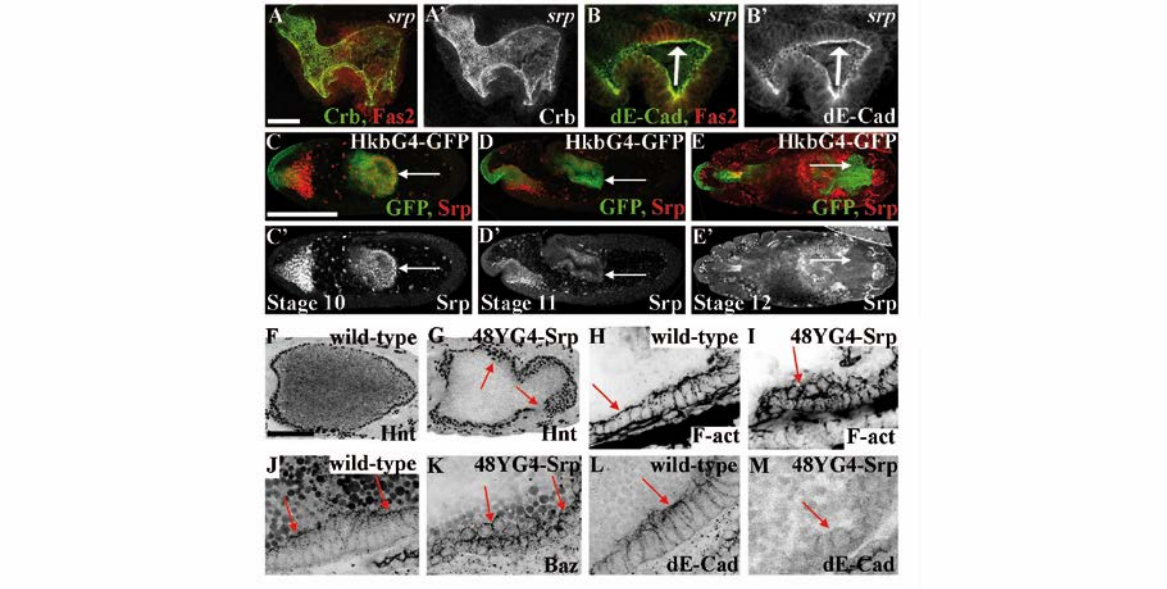



**Figure 2. Srp is required for PMG EMT and maintenance of Srp expression prevents PMG cells from redeveloping epithelial characteristics.** A,B, stage 12 *srp* mutants stained for the basolateral marker Fascicilin2 (Fas2), and Crb (A) or dE-Cad (B). In *srp* mutants, PMG cells remain columnar in shape (B, arrow) and tightly localise both Crb and dE-Cad to their apical domains. C-E, Srp expression during stages 10 (C), 11 (D) and 12 (E), arrows point to the PMG. Srp expression is lost from PMG cells during stages 11-12 (D,E). F,H,J,L stage 15 wild-type and g,i,k,m sustained expression of Srp in the endoderm using 48Y-Gal4. (F,G) Srp overexpression leads to a discontinuous midgut epithelium, containing breaks and cell multi-layering (compare F with G, arrows). Both F-act and Baz are found delocalised around the cell cortex (I,K arrows), and dE-Cad is almost undetectable in the cells (M, arrow). Black and white fluorescence images were colour-inverted in Photoshop. Scale bars: A,B, 20 μm; C-E, 100 μm; F,G, 50 μm; H-M, 20 μm. See also Supplementary Figure S2; page 101.

## Ectopic expression of Srp induces a similar EMT in epithelial cells

To further explore the ability of Srp to promote such transitions, *srp* was ectopically expressed in stripes of cells in the embryonic epidermis, which

remains epithelial throughout development. Upon *srp* overexpression, cells appear disorganised and dispersed, with stripes broken and fusing across compartment boundaries (Figure 3B, Supplementary Figure S3B; page 102). Furthermore, the cells are no longer present as a monolayer at the cell surface and instead, form multilayered clusters projecting into the embryo (Figure 3D-F, arrows). Cells lose their columnar morphology and become more elongated and irregular in shape (Figure 3D-F). Internalised epidermal cells localise F-Act throughout the cell cortex and do not express Crb, and dE-Cad becomes relocalized around the cell (Figure 3E,F, Supplementary Figure S3D,E; page 102). In some internalised cells dE-Cad expression is below detectable levels (Figure 3E, arrow). While it could be that dE-Cad is harder to detect in these cells becauseof its delocalisation, it suggests that very high levels of *srp* may induce a downregulation/degradation of dE-Cad protein. Interestingly, the few *srp* overexpressing cells that remain at the surface of the embryo express high levels of Crb and dE-Cad at the apical surface, but have undergone apical constriction (Figure 3D,F,G), a common feature of developmental EMTs (Shook and Keller, 2003).

Analysis in living embryos shows that *srp* overexpression leads to the activation of migratory behaviour, with cells extending many protrusions (Supplementary Movies S3 and S4) and moving to more distant sites in the embryo (Figure 3I-L, arrowheads in J and L indicate long cords/groups of cells mispositioned in the embryo, Supplementary Movies S5 and S6). Notably, while cells appear more elongated and protrusive, they do not become dispersed but move collectively through the embryo. Furthermore, *srp* overexpression leads to induction of Matrix metalloproteinase 1 (Mmp1) (Figure 3G), which is known to cleave extracellular matrix components, and thus facilitate invasive migratory cell behaviour (Srivastava et al., 2007). Ectopic Srp also induces expression of Forkhead (Fkh) (Figure 3H), an orthologue of the vertebrate FoxA transcription factors. *fkh* has previously been reported to act as a survival factor in a number of *Drosophila* systems, including the midgut (Cao et al., 2007; Myat and Andrew, 2000; Tepass et al., 1994), suggesting that similar to vertebrate *sna* (Barrallo-Gimeno and Nieto, 2005), *srp* overexpression may cause cells to become more resistant to apoptosis and cell death. Altogether, these results indicate that *srp* expression in epithelial cells triggers a loss of epithelial characteristics and acquisition of mesenchymal features (Figure 3M).

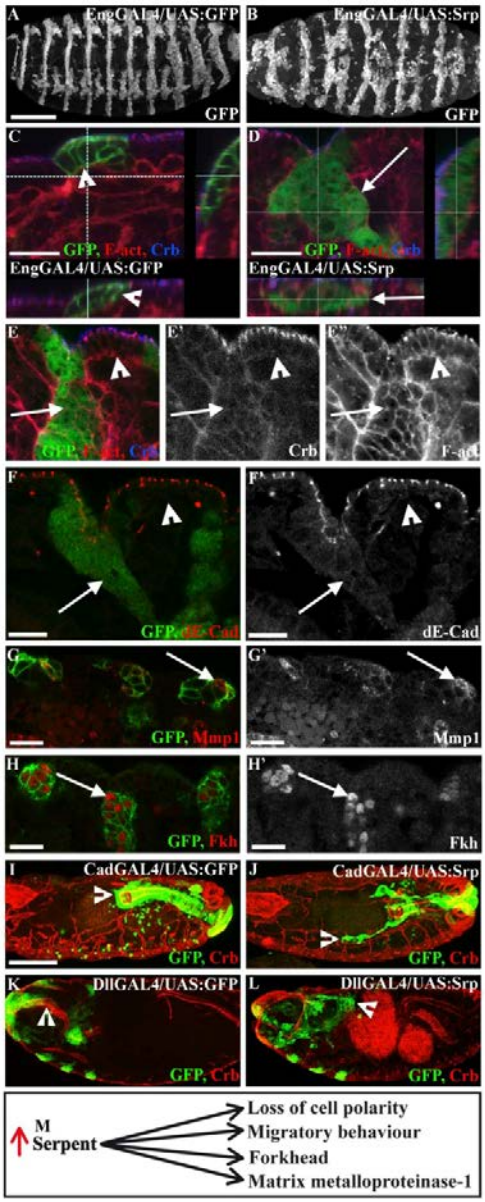



**Figure 3. Ectopic expression of Srp induces loss of epithelial traits and acquisition of migratory and invasive behaviour.** A-H stage 16 embryos expressing either GFP alone under control of EngGal4 (A,C) or both Srp and GFP (B,D-H). Epidermal cells normally form a monolayer at the cell surface (C,E,F, arrowheads). Srp expressing epidermal cells are found in multilayered clusters that project into the embryo (D-F, arrows, compare with wild-type C). In internalised cells F-act is found all around the cell cortex (D, arrow), Crb (D) and dE-Cad (E) are reduced to low levels, or are undetectable (D-F, arrows). Expression of Mmp1 (G) and of Fkh (H) is induced in Srp positive epidermal cells. I-L stage 16 embryos expressing either GFP alone (I,K) or both Srp and GFP, (J,L) in epithelia in the posterior of the embryo, using Cad-Gal4, which is expressed in the hindgut and anal pads, (I,J), or in the anterior, using Dll-Gal4 (K,L). Srp overexpressing epithelia are found at more distant sites in the embryo to in wild-type (compare arrowheads in I,K with J,L). Individual green cells in I and J are haemocytes, which have phagocytosed GFP positive apoptotic cells. M, Schematic summarising the effects of Srp overexpression. EngGal4: Engrailed-Gal4, drives in posterior compartments of the ectoderm; CadGal4: Caudal-Gal4, drives in the posterior of the embryo; DllGal4: Distal-less-Gal4, drives in the anterior of the embryo and some imaginal discs. Scale bars: A,B, 100 μm; C-H, 10 μm; I-L, 100 μm. See also Supplementary Figure S3; page 101 and Supplementary Movies S3, S4, S5 and S6.

## Srp-induced epithelial transitions do not act through a repression of dE-Cad transcription

The transition from epithelial to migratory non-polarised cell states involves coordinated changes in numerous cell features such as polarity, shape, adhesion and migratory and invasive capacities, which in turn are affected by transcription factors such as Sna through the activation and repression of many target genes (reviewed in (Barrallo-Gimeno and Nieto, 2005), and this is likely to also be the case for Srp-induced EMT. However, it is clear that while EMT involves many changes in cell behaviour, a key step in the initiation of EMT is a functional loss of E-Cad and thus of cell-cell adhesion (Huber et al., 2005); (Peinado et al., 2007). Hence we decided to investigate if Srp effects adhesion through the repression of E-Cad transcription, as has been found for almost all EMT regulators reported to date (Thiery et al., 2009). During endodermal-EMT dE-Cad loses its apical localisation and there is a marked decrease in junctional protein levels (Figure 1P-R). However, whole mount *in situ* hybridization shows that dE-Cad RNA remains expressed in the PMG during endodermal-EMT, and is expressed in migrating PMG cells (Figure 4A-C). aPKC, which behaves similarly to dE-Cad during PMG-EMT and migration, also shows continued RNA expression (Supplementary Figure S4A-C; page 103). Next, we quantified the levels of *dE-Cad* RNA in the PMG at time-points before and after endodermal-EMT (see Experimental Procedures) and found that there are no significant variations in *dE-Cad* levels between these two stages (Figure 4D). These data suggest that in contrast to *sna*, *srp* does not act through the repression of dE-Cad transcription, and this is reinforced by the following three observations. Firstly, sustained expression of *srp* in the midgut does not lead to a loss of dE-Cad RNA, in comparison to *sna* overexpression, which causes dE-Cad RNA to be almost completely absent from midgut cells (Figure 4D-I). Secondly, *srp* overexpression does not lead to the induction of known repressors of *dE-Cad* such as *sna* and the ZEB related factors *zfh1* and *zfh2* (Supplementary Figure S5; page 104). Finally, the overexpression of *sna* in the ectoderm does not induce EMT; this is likely due to overexpression of *sna* repressing only the zygotic *dE-Cad* transcription, while the ectoderm is a tissue where the maternal contribution of RNA is sufficient to maintain the tissue throughout embryogenesis (Tepass et al., 1996; Uemura et al., 1996). We conclude that Srp does not induce endodermal-EMT through the repression of *dE-Cad* transcription, rather it triggers a downregulation of junctional dE-Cad protein through the activation/repression of downstream targets.

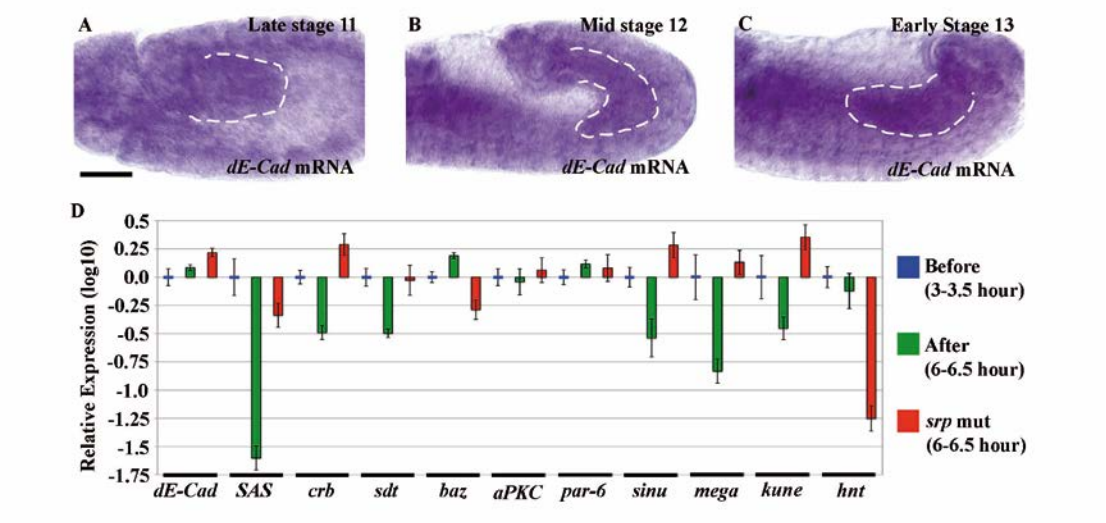



**Figure 4. crb, sdt and SAS are transcriptionally repressed during PMG-EMT, while dE-Cad and components of the Par3 complex continue to be transcribed.** A-C *in situ* hybridization in wild-type embryos for *dE-Cad*. *dE-Cad* remains strongly expressed in the PMG throughout stages 11 (A), 12 (B) and 13 (C). (D) qRT-PCR analysis of *dE-Cad, SAS, crb, sdt, baz, aPKC, par-6, sinu, mega, kune* and *hnt* RNA levels in PMGs dissected from either 3-3.5 hour (Before), 6-6.5 hour old wild-type embryos (After), or from 6-6.5 hour *srp* mutant embryos (*srp* mut). There are no significant changes in the RNA levels of *dE-Cad, baz, aPKC* and *par-6* in the PMG between 3 and 6.5 hours of development, or between wild-type and *srp* mutants. In contrast, *SAS, crb*, *sdt, sinu, mega and kune* become strongly repressed, and this repression is lost in *srp* mutants. *hnt* was used as a control, as it is a known downstream target of Srp, and does not change in expression in the PMG between 3 and 6 hours of development. Gene expression levels were normalized using the endogenous control Actin5C. Error bars indicate SD. Forward and reverse primers are provided in Supplementary Experimental Procedures. See also Supplementary Figures S4 and S5; pages 103-104.

## Srp directly represses *crb* transcription

How does Srp induce a loss of cell polarity and adhesion? To address this question we screened components of the different cell-membrane domains using qPCR to identify genes that were expressed in the PMG prior to endodermal-EMT, but which become repressed in the PMG in a Srp-dependent manner (see Experimental Procedures). We found that genes encoding the apically localised proteins SAS, Crb and Sdt, and the basolaterally localised Claudin proteins, Sinuous (Sinu), Megatrachea (Mega) and Kune-kune (Kune) become strongly repressed in the PMG during endodermal-EMT, and that this repression is lost in *srp* mutants (Figure 4D). In contrast, we found that there are no significant changes in the levels of the apical Par genes *baz* and *par-6*, or of *aPKC*, during endodermal-EMT, or in *srp* mutants (Figure 4D).

We decided to focus further studies on the key cell polarity regulator Crb, as *crb* is required for the maintenance of epithelial cell polarity and stabilisation of adherens junctions (Tepass et al., 1990); (Grawe et al., 1996), and in the

ectoderm of *crb* mutants dE-Cad is lost from the apical domain and cells lose cell-cell adhesion (Tepass et al., 1990); (Grawe et al., 1996), similar to what happens during PMG-EMT. Whole mount *in situ*s confirmed that *crb* is highly expressed and apically localised in the hindgut and PMG cells prior to stage 10 (Figure 5A), and lost from the PMG during endodermal-EMT, so that its expression is almost undetectable by whole mount *in situ* hybridization in late stage 11 and 13 PMG cells (Figure 5B,C). *In situ* hybridization studies in *srp* mutants showed that *crb* RNA levels remain high in PMG cells throughout embryogenesis (Figure 5D), reinforcing the quantitative analysis, and further confirming that *crb* expression is directly or indirectly regulated by Srp.

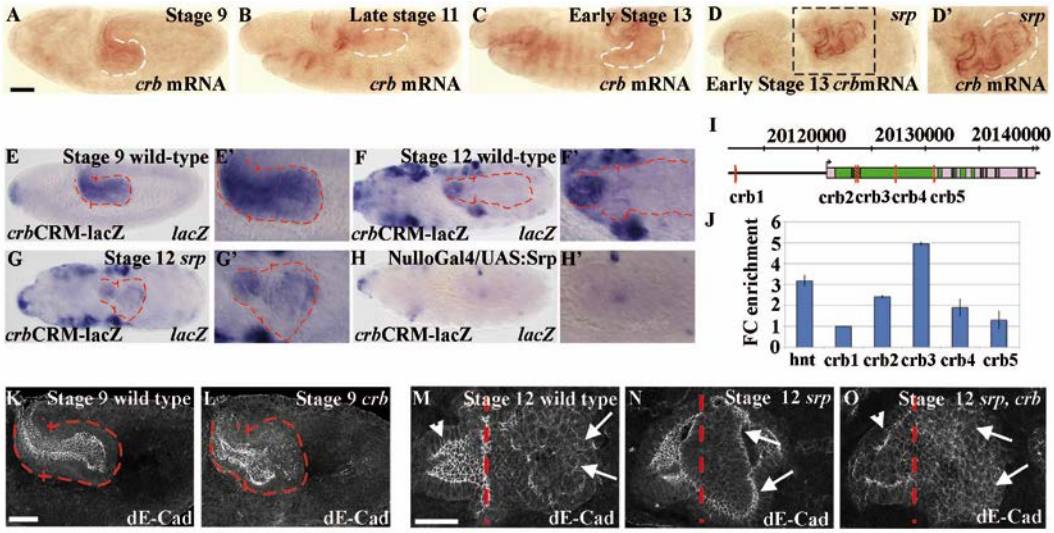



**Figure 5. Srp represses crb transcription.** A-C *in situ* hybridization for *crb* in wild-type embryos. *crb* is expressed in the endoderm throughout stage 9 (A) but is not detected in the PMG after it undergoes EMT (B,C). (D) *crb* is strongly expressed in the PMG in stage 13 *srp* mutants. (E,F) The *crb*CRM-lacZ reporter reproduces the endogenous crb expression pattern in the PMG (compare E,F with A,C). (G) In *srp* mutants, the *crb*CRM-lacZ reporter is not repressed in the PMG (compare G with wild-type F) (H) Overexpression of Srp throughout the embryo, using Nullo-Gal4, causes a general repression of the *crb*CRM-lacZ reporter (compare H with F, arrowheads) (I) schematic diagram indicating the location of putative Srp-binding sites (red lines) with respect to the *crb* gene (exons in purple, introns in green), and the location of the *crb*-CRM (shaded in black). (J) ChIP assays, performed with Srp antibody; data are expressed as fold change enrichment of Srp-binding relative to the negative control, and are normalized to input. Forward and reverse primers are provided in Supplementary Experimental Procedures. Error bars indicate SD. (K,L) In *crb* mutants dE-Cad becomes delocalised in PMG cells during stage 9 (compare L with wild-type K). M,N,O Stage 12 wild-type (M), *srp* (N) and *srp, crb* double mutant embryos (O) stained for dE-Cad, red lines show the hindgut PMG boundary. (N) In *srp* mutants PMG cells do not undergo EMT and dE-Cad is apically localised throughout the hindgut (arrowhead) and PMG (arrows). (O) in *srp ,crb* double mutants, dE-Cad becomes delocalised in PMG cells (arrows in O), similar to in wild-type (arrows in M) and in contrast to in *srp* mutants (arrows in N), while the apical localisation of dE-Cad in the hindgut is retained (arrowhead in O). Dotted lines outline the PMG, vertical lines show the hindgut PMG boundary; black box shows the area magnified in D'. Scale bars: A-H, 50 μm; J-N, 20 μm.

To further examine the relationship between Srp and *crb,* we performed a ChIP analysis to determine whether Srp is physically associated with *crb* during early stages of development. We focused on a region situated within the 1st intron of *crb*, termed the *cis*-regulatory module (CRM), as this region has previously been shown to reproduce the endogenous *crb* expression pattern in the mesoderm (Sandmann et al., 2007), and we found that this is also true for the PMG (Figure 5E,F). We identified many putative Srp-binding sites within the 1st intron of *crb,* and tested for Srp binding to these sites by ChIP-PCR analysis. As a positive control for this experiment we found that a putative Srp-binding site in the 1st intron of *hindsight* (*hnt*), a known target of Srp (Yip et al., 1997), is positive for Srp-binding (Figure 5I,J). We found that the two sites located within the *crb*-CRM, Crb-2 and Crb-3, show a similar or greater fold change enrichment when compared to hnt, and thus are positive for Srp-binding (Figure 5I,J), pointing to a direct interaction between Srp and *crb* through the *crb*-CRM regulatory region. Conversely, other putative Srp-binding sites outside the CRM were negative for the same assay (Figure 5I,J and data not shown) indicating its specificity. To further test if Srp acts through the *crb*-CRM to repress *crb* expression we next examined the behaviour of the *crb*-CRM-lacZ reporter in *srp* mutant and *srp*-overexpressing backgrounds. We found that when Srp is absent, the reporter remains highly expressed in the PMG throughout development (Figure 5G) and conversely when Srp is overexpressed throughout the embryo, we see that expression of the reporter is almost completely lost (Figure 5H). Taken together these results suggest that Srp directly represses *crb* through binding to GATA sites within the *crb*-CRM region.

**Loss of Crb partially reverts the dE-Cad phenotype in a Srp mutant**

To investigate the functional relevance of repression of *crb* by Srp we first examined the behaviour of dE-Cad in the PMG of *crb* mutants. We found that in *crb* mutants the PMG undergoes a premature delocalisation of dE-Cad, with dE-Cad becoming lost from PMG cells as early as stage 9, and a premature alteration of epithelial morphology (Figure 5K,L). To provide a functional validation of *crb* repression in the Srp-induced epithelial transition, we generated a double mutant for *srp* and *crb* and found that loss of *crb* partially reverts the dE-cad phenotype and the failure of EMT in the PMG caused by *srp* mutants (Figure 5M-O, arrows). It should be noted that while dE-Cad becomes delocalised in the PMG of *srp; crb* double mutants, it remains tightly localised in the hindgut (Figure 5O, arrowhead). Together these data indicate that Srp induces a delocalisation of dE-Cad through the direct repression of *crb.* Thus the repression of *crb* by Srp appears to be a central part of the mechanism used to induce endodermal-EMT. However, we have found that ectopic expression of *crb* does not inhibit PMG morphogenesis (data not shown). While this could be

due to ectopic Crb not localising correctly to the apical membrane, it indicates that *crb* downregulation is necessary for endodermal-EMT but *crb* expression is not sufficient to abrogate it. This is in agreement with other genes such as *std* and *SAS* being repressed by Srp in the PMG. Furthermore it is similar to results for Sna-induced EMT where despite the fact that repression E-Cad is a crucial step in this process (Cano et al., 2000); (Batlle et al., 2000); reviewed in (Huber et al., 2005), overexpression of E-Cad alone is not sufficient to block EMT (Ohkubo and Ozawa, 2004).

**Vertebrate GATA6 induces a similar EMT in MDCK cells**

Vertebrate orthologues of Srp, GATA 4, 5, and 6 (Gillis et al., 2008) are similarly expressed and implicated in the differentiation of various organs of endodermal origin such as intestine, colon and liver (Gao et al., 1998); (Molkentin, 2000). These genes appear to have distinct roles in mammalian development and pathogenesis. In particular, hGATA6 has recently been reported to be upregulated in some human cancers of endoderm origin (Shureiqi et al., 2007); (Fu et al., 2008); (Kwei et al., 2008), and an increased accumulation of hGATA6 has also been found in cells at the leading front of some endodermal tumors (Haveri et al., 2008). To study the putative role of hGATA6 in EMT, we engineered MDCK cells to inducibly express hGATA6. This epithelial cell line has previously been instrumental in discovering the role of Snail, Slug and Twist in EMT (Cano et al., 2000); (Yang et al., 2004); (Bolos et al., 2003). Expression of hGATA6 in MDCK cells caused a dramatic conversion towards a spindle mesenchymal-like phenotype as opposed to the epithelial morphology of non-induced control cells (Figure 6B,E). This morphological change was concomitant with a loss of membrane-associated E-Cad, which accumulated in punctae in the cell cytoplasm (Figure 6C,F), but does not change in overall protein levels (Figure 6I). hGATA6 also induced remodelling of F-Act from the cell cortex to stress fibres as well as to motile structures similar to lamellipodia (Figure 6C',F'). Furthermore, hGATA6-expressing cells displayed membranous N-Cadherin (N-Cad) (Figure 6G,H), a well-established marker gene of mesenchymal cell (Sobrado et al., 2009). To test for changes in migratory and invasive capacity in hGATA6 expressing cells, we performed assays on collagen type IV-coated Transwells. We found that hGATA6 expression greatly enhances the motility of MDCK cells, in comparison to non-induced control cells (Figure 6J-L). Together these results show that hGATA6 triggers a loss of epithelial traits, acquisition of fibroblastic morphology and increases motility and invasive capacity in MDCK cells.

## Srp and GATA6-induced transitions share some features

To further explore the similarities between Srp and hGATA6-induced transitions we next investigated for changes in the RNA levels of *E-cad* upon expression of hGATA6. We found that similar to Srp, hGATA6 does not repress *e-cad* transcription, as there are no significant differences in *e-cad* RNA levels between hGATA6 expressing MDCK cells and non-inducible controls (Figure 6M). Thus both Srp and hGATA6-induced transitions occur without abolishing E-Cad transcription.

We next examined for changes in the RNA levels of members of the Crumbs complex upon expression of hGATA6. There are three vertebrate Crumbs genes - *crb1*, *crb2* and *crb3* and of these, just *crb2* and *crb3* are expressed in MDCK cells (Roh and Margolis, 2003). We found that while *crb3* RNA expression levels do not change upon hGATA6 expression, *crb2*, is strongly downregulated (Figure 6M). Other members of the Crb complex, *protein associated with Lin-7 (pals1)* or *Pals1-associated tight junction protein (PATJ)*, did not show any changes at the RNA level (Figure 6M). 4. Next, we made use of a published data set of hGATA6 binding sites that was generated from an hGATA6 RNA-seq experiment performed on the human intestinal Caco-2 cell line (Verzi et al., 2010). We searched their processed data for positive hGATA6 binding peaks in the regulatory regions of crb2 and crb3 and found that while there are no peaks in *crb3, crb2* has a positive GATA6-binding peak within its first intron (Verzi et al., 2010). Genomic alignment shows that the hGATA6 binding region is conserved from humans to dogs (data not shown), further supporting the idea that similar to Srp in flies, hGATA6 represses *crb2* in MDCK cells through direct binding to a regulatory region within the first intron of *crb2*.
Finally we assayed for changes in the expression of Claudin genes, and found that *claudin 1, 7* and *16* are all strongly repressed by hGATA6 (Figure 6M), while expression of hGATA6 induces an upregulation of *mmp1* (Figure 6M). Thus, the repression of Claudin genes and activation of *mmp1* are also a conserved feature of Srp and hGATA6-induced transitions.

## Vertebrate hGATA4 also induces an EMT in MDCK cells

As mentioned above, other vertebrate GATA factors such as GATA4 and 5 are also closely related to Srp. Thus, we performed with hGATA4 and 5 similar experiments in MDCK cells as those performed with hGATA6 and found that hGATA4 has a similar effect to hGATA6 (Supplementary Figure S6G,H; page 105). However, EMT was not induced by hGATA5 (Supplementary Figure S6E,F; page 105). This indicates the specificity of the transformation, which appears not to be a common feature of all GATA factors, and is in agreement

with the observation than the vertebrate GATA4 and 6 are both more closely related to each other than to the GATA5 gene (Gillis et al., 2008; Lowry and Atchley, 2000). Furthermore GATA4 and 6 have recently been reported to have overlapping functions in the regulation of proliferation and differentiation of intestinal cells (Beuling et al., 2011). As a control, we also induced the expression of a mutated form of hGATA6 (which contains point mutations in both zinc fingers and thus cannot bind to DNA) and found that it did not induce an EMT (Figure 6I,J). Although it is still not clear which are the mechanisms involved in hGATA6 induced E-Cad protein relocalization, altogether, these results demonstrate that among some GATA factors, the ability to induce a distinct EMT without silencing E-cad transcription is a conserved feature from flies to humans.

# DISCUSSION

In this study we investigate the role of Srp in the *Drosophila* endoderm, and show that it is required to induce endoderm cells to switch from an epithelial to non- polarised motile behaviour, and that it acts as a potent trigger of a similar transition when miss-expressed spatially or temporally in epithelial cells. Srp promotes dE-Cad junctional downregulation and while Srp is likely to impinge on a large number of genes, a specific feature of its activity is *crb* transcriptional repression. In addition, we have also shown that the overexpression of either hGATA6 or 4 in MDCK cells triggers a distinct EMT, which shows many similarities to the Srp-induced transition, as changes in cell polarity, adhesion and motility occur without any changes in E-Cad at the RNA or protein levels, while a crb orthologue, *crb2* becomes strongly repressed. Cells traditionally have been classified as epithelial or mesenchymal depending upon whether they are polarized or not, possess junctions or have lost adhesion, and display static or migratory and invasive behaviour. Here we show that during the formation of the endoderm, cells lose apico-basal polarity and become highly motile, a common feature of both developmental and pathological EMTs. However, a distinct feature of the endodermal transition is that while adherens junctions become fragmented, dynamic puntae of dE-Cad protein are maintained at the plasma membrane. Despite their highly mesenchymal appearance, endoderm cells migrate as a collective mass. Thus, spots of dE-Cad may be retained at the membrane during this type of transition to facilitate collective, rather than individual migration. It should be noted, this is very different to other types of collective migration which have described for epithelial cells in the *Drosophila* embryo, such as in the trachea, where cells migrate, but do not lose apico-basal polarity or disassemble cell junctions.

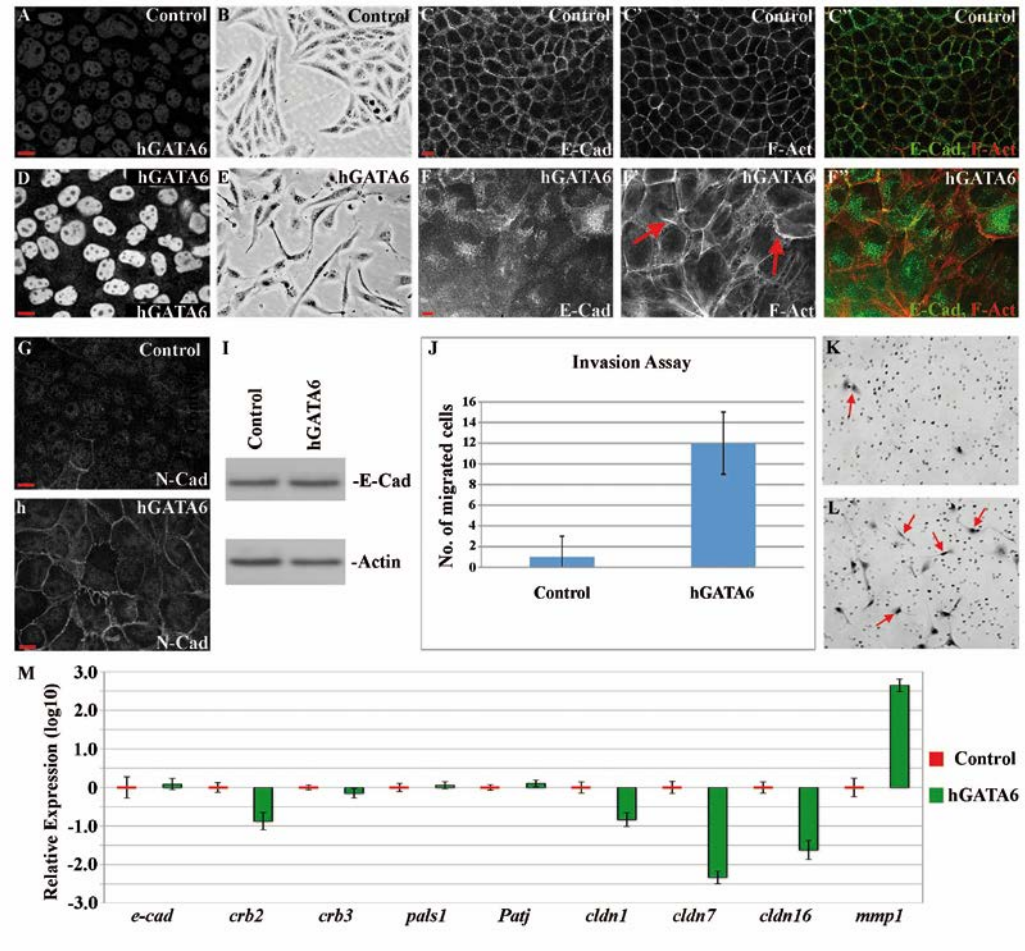



**Figure 6. hGATA6 induces EMT in MDCK cells.** A-H MDCK-hGATA6-dox cells before (A,B,C,G) and 6 days after induction of hGATA6 expression (D,E,F,H). (A,D) Nuclear accumulation of hGata6 protein after addition of doxycycline. (B,E) hGATA6-induced MDCK cells adopt a mesenchymal-like spindle shaped morphology. (C,F) Single confocal slice of cells stained for F-Act and E-Cad. The normal junctional localisation of E-Cad is lost, and instead E-Cad is found in punctate dots in the cell cytoplasm. F-Act appears reorganised into stress fibres and lamellipodia-like structures (F', arrows). (G,H) N-Cad is found localised at the cell membrane of hGATA6-induced cells. (I) Western blot analysis show that hGATA6 expression does not induce changes in the levels of E-Cad protein. (J) The migratory and invasion capacity of control and hGATA6-induced MDCK cells was analysed in transwell assays on matrigel 96 hours after hGata6 induction. hGATA6-induced cells possessed a 10-fold increase in migration/invasion behaviour. Eight randomly selected microscopic fields (20X magnification) were imaged to calculate the average number of occupied pores per microscopic field. Results represent the mean ± S.D. of at least two independent experiments performed in duplicate. (K,L) Cells migrated through the matrigel to the bottom (K,L arrows). (M) qPCR analysis shows the relative expression changes, in log10 scale, of the indicated genes in hGATA6 expressing MDCK cells (green) compared to the non-induced controls (red) 48 hours after doxycycline induction. Gene expression levels were normalized using the endogenous control GAPDH. Error bars indicate SD. Scale bars, 10 μm. See also Supplementary Figure S6; page 105.

EMT in the mesoderm and almost all EMTs studied to date, have been shown to rely on the activity of *sna* genes (Barrallo-Gimeno and Nieto, 2005). The few other cases that do not depend on *sna* genes, rely on the activation of other E-cad transcriptional repressors such as E47 (Thiery et al., 2009). In contrast, the transition reported here occurs independently of Sna and does not rely on E-Cad transcriptional repression. Thus, the endodermal-transition is distinct from canonical EMTs not only at the morphological, but also at the transcriptional level. One way of achieving different degrees of adhesion between cells subsequent to EMT could be to drive Sna at different levels, however this could be more difficult to achieve and more prone to errors than activating an independent mechanism. Thus rather than activating a single EMT mechanism at different intensities, Srp appears to be a novel alternative trigger for this distinct kind of EMT in which low levels of dE-Cad are retained at the membrane.

We and others have reported many cases where E-Cad is downregulated at the protein level during development (i.e. (Shaye et al., 2008)). This also occurs during EMT in gastrulating mouse embryos, where p38 downregulates E-Cad at the protein level by a mechanism independent of transcriptional repression by Sna (Zohn et al., 2006). However, what is different about the Srp-induced epithelial transition reported here, is that while it is also not linked to *dE-cad* gene repression, it is both necessary and sufficient to induce a transition to non-polarised, motile cell behaviour. This contrasts with other reported cases, such as in the *Drosophila* trachea, where dE-Cad is downregulated, but this is not sufficient to induce a loss of cell polarity. In the case of mouse gastrulation, *sna* mutant cells fail to undergo EMT (Carver et al., 2001); (Barrallo-Gimeno and Nieto, 2005), indicating that the p38-dependent E-cad delocalisation is not sufficient to drive EMT and in p38 mutants there is still some EMT (Zohn et al., 2006), indicating that the p38 pathway is not absolutely required for EMT. This strongly contrasts with the Srp pathway, which is absolutely required and sufficient to drive a transition to non-polarised migratory cell behaviour.

The specific features of the Srp-induced EMT are likely to facilitate cohesive cell migration, and it seems easier to revert to epithelial again, a case often found both in development and tumor progression. Indeed, the collective invasion of cells that display combination of epithelial and mesenchymal features rather than a full conversion to a mesenchymal phenotype is prevalent in many cancer types (Friedl and Gilmour, 2009). The finding that GATA factors play a conserved role in inducing a similar epithelial transition in mammalian cells and that high levels of expression of hGATA6 are found in a number of endodermal tumors highlight the relevance of our results for development and cancer research. Further studies will be required to assess whether hGATA6

expression in such tumors induce similar changes in cell behaviour to those seen in flies and MDCK cells.

## ACKNOWLEDGEMENTS

We are grateful to members of the Casanova and Llimargas labs and Angela Nieto for helpful discussions and to Nicolás Martín, Yolanda Rivera and members of the Jiménez lab for technical assistance. We also acknowledge the Advanced Digital Microscopy, the Biostatistics/Bioinformatics Unit and the Functional Genomics Core Facilities from the Institute of Research in Biomedicine and the Bioinformatics Platform from the Consolider Project GESHAPE, CSD 2007-00008. We thank E. Furlong, D. Hoshizaki, E. Knust, J. Kumar, H. Oda, B. Sanson, S. Sotillos, A. Wodarz, the Bloomington stock centre and DSHB for kindly sending us flies and reagents. This work has been supported by the Generalitat de Catalunya, the Spanish Ministerio de Ciencia e Innovacion and its Consolider-Ingenio 2010 program. We also acknowledge support from the Programa Juan de la Cierva (K.C), (E.B) (G.W) and (X.F.M.)

## EXPERIMENTAL PROCEDURES

### Immunohistochemistry, *in situ* hybridization, image acquisition and analysis

Embryos were fixed, mounted and staged using standard techniques. Embryos were stages according to the staging scheme in (Hartenstein V., 1993). To determine exactly when EMT takes place in the PMG, embryos were collected over 30 minute time periods, and then analysed for markers for EMT. Antibodies and probes used are detailed in the Supplementary Experimental Procedures. Confocal images were acquired with a Leica SP5. Images were post-processed with Adobe Photoshop and ImageJ.

### Chromatin Immunoprecipitation (ChIP)

ChIP experiments were carried out as described previously (Sandmann et al., 2006). In each ChIP assay 4 independent chromatin samples were prepared from 0-6 hour wild-type embryo collections, 2 samples were incubated with anti-Srp sera and 2 samples without antibody, as a negative control. Each assay was preformed 3 times, and only regions that showed a strong enrichment in all 3 assays were considered to be positive for Srp-binding. Primers used to assay for regions containing putative Srp-binding sites are listed in the Supplementary

Experimental Procedures.

**Quantification of RNA levels in the PMG**

The PMG was dissected from wild-type embryos at 3-3.5 hours (Before), 6-6.5 hours (After) or from *srp* mutant embryos at 6-6.5 hours, according to the method described in (Skaer, 1989). 3 PMGs were dissected per condition, RNA was isolated from the cells and amplified by Pico Profiling (Gonzalez-Roca et al.). 4 separate sample sets were isolated, and the RNA amounts quantified by quantitative real-time PCR. Primers used are detailed in the Supplementary Experimental Procedures. Gene expression levels were normalized using the endogenous control Actin5C for each sample, and differences in target gene expression were determined using the StepOne 2.2 software.

**Immunofluorescence of MDCK cells**

MDCK cells were plated at a high confluence (150 000 cells/mL) and grown for 48h on sterile transparent 0.4µm pore polyester membrane inserts (Corning 3470) coated with 2.5 µg/cm$^2$ laminin (Sigma, L2020). For full details of fixing and staining procedures, and antibodies used see Supplementary Experimental Procedures.

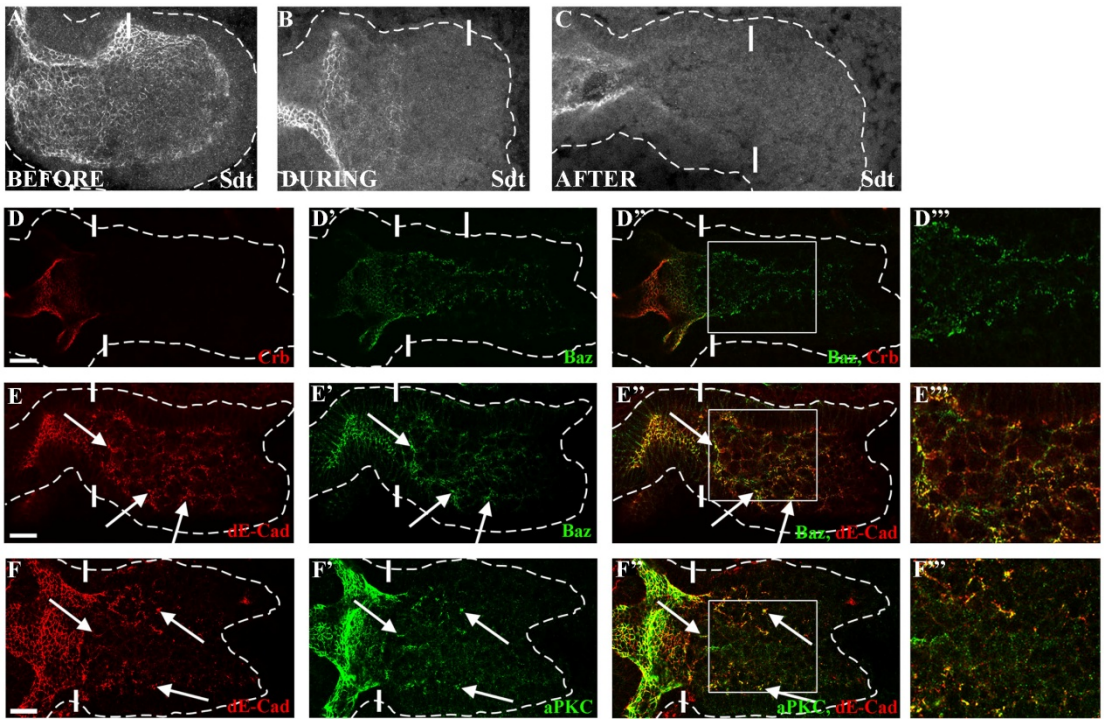
**Invasion Assay**

To obtain polarized monolayers for immunofluorescence, MDCK cells were seeded on transparent track-etched PET chambers, coated with matrigel. hGATA6 and control cells were seeded at a density of 5000 cells/well in DMEM medium supplemented with 0.5% BSA. Lower chambers were filled with 10% FBS to act as a chemoattractant. After 48 h of incubation, the cells on the lower membrane surface were fixed and stained.

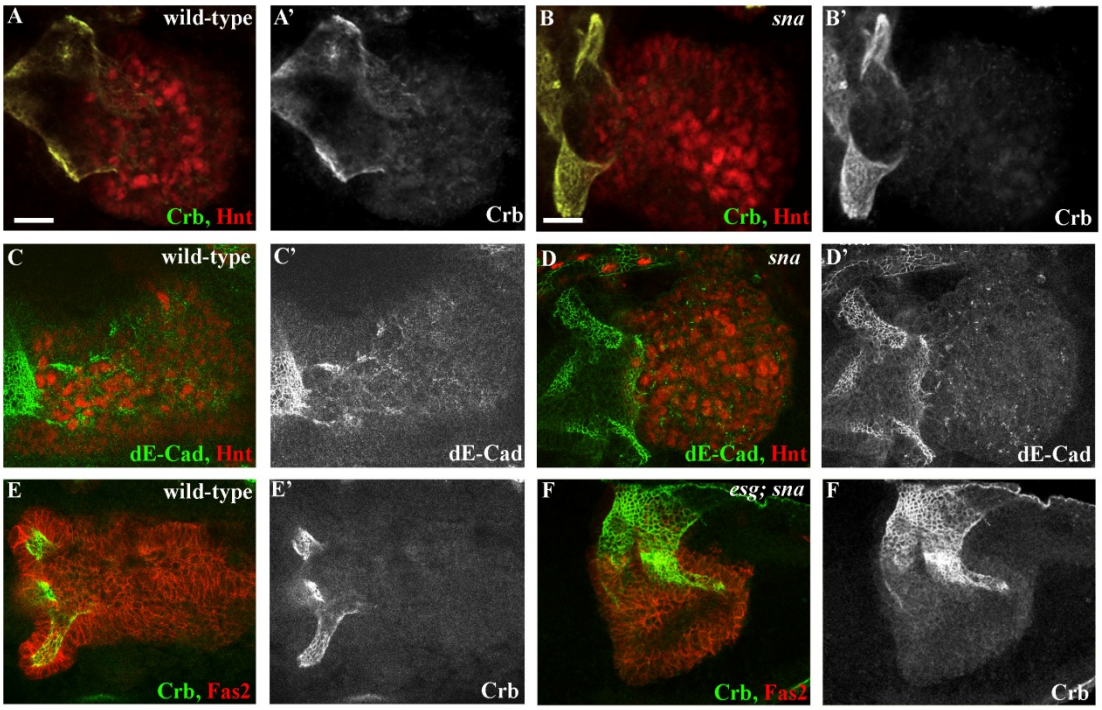
**Quantitative Real Time-PCR on MDCK cells**

Total RNA samples were extracted using TRIzol, DNAase treated, further purified using RNeasy columns (Quiagen) and reversed transcribed using the High Capacity cDNA Archive Kit (Applied Biosystems). RNA amounts were quantified by quantitative real-time PCR. Primers used are detailed in the Supplementary Experimental Procedures. Gene expression levels were normalized using the endogenous control GAPDH for each sample, and differences in target gene expression were determined using the StepOne 2.2 software.

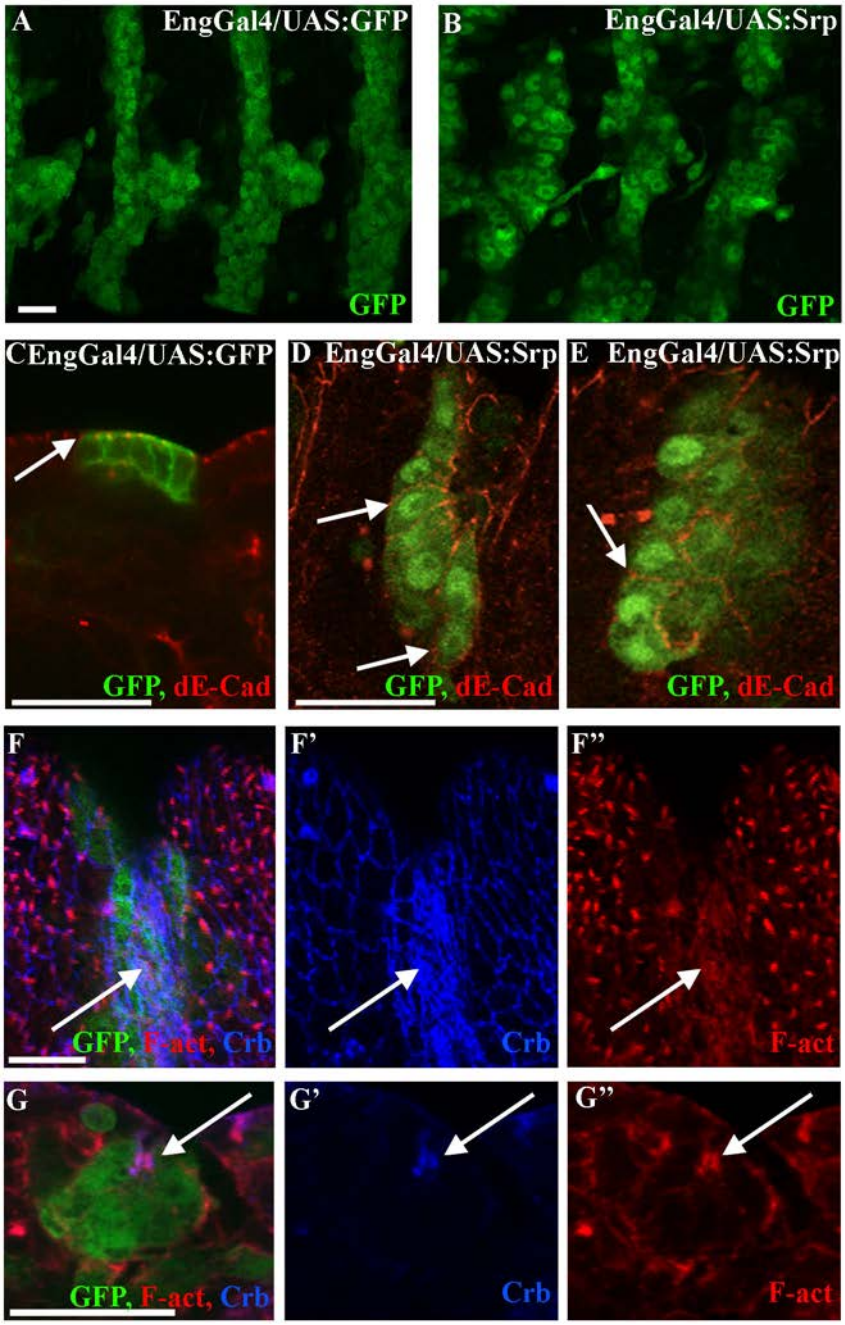
# SUPPLEMENTARY FIGURES



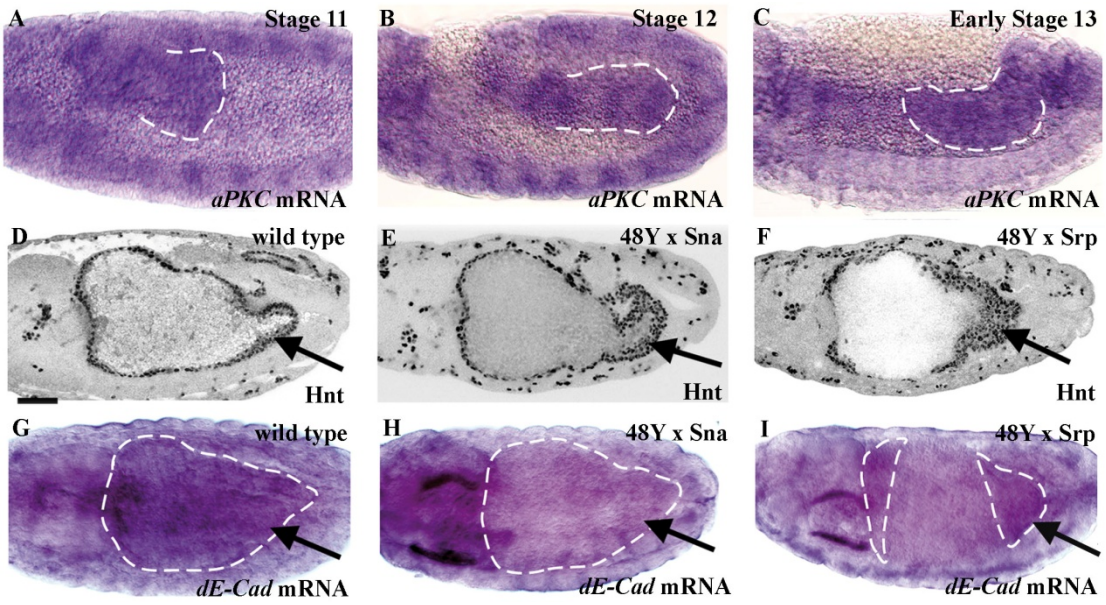

**Supplementary Figure S1. Sdt and Crb are lost from PMG cells as they initiate migration, while punctate spots of of Baz, aPKC and dE-Cad remain.** A-C stage 9 (A), stage 10 (B) and early stage 12 (C) wild-type embryos stained for Sdt. D-F stage 11 wild type embryos stained for Baz and Crb (A) , Baz and dE-Cad (B) or aPKC and dE-Cad (C). a, Crb in undetectable in stage 11 PMG cells. B,C Baz, dE-Cad and aPKC co-localise to punctate spots (B,C arrows). White boxed regions show the areas magnified in A'''-C'''. Scale bars, 20 μm.



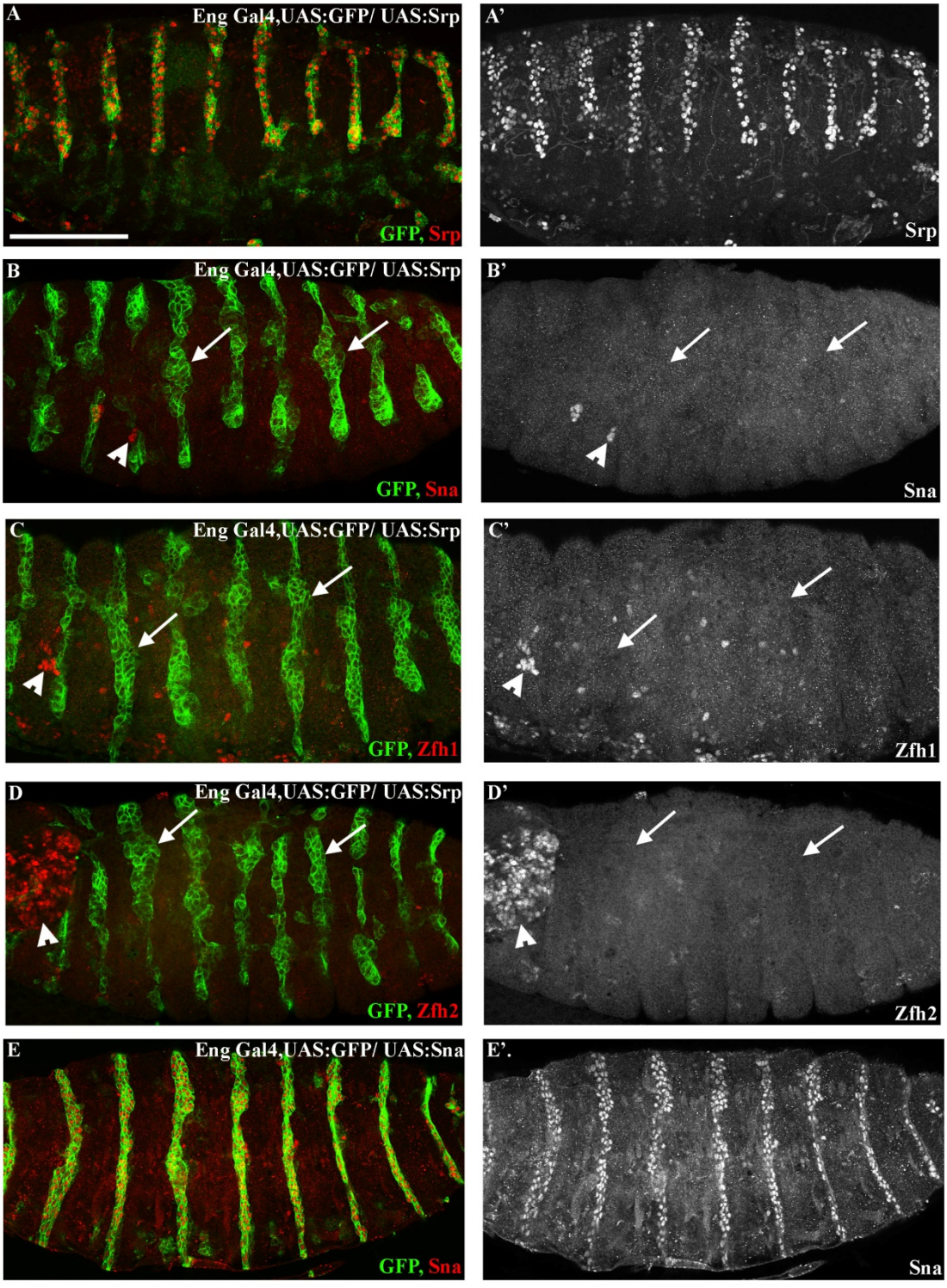

**Supplementary Figure S2. Sna and Esg are not required for EMT in the PMG.** A-F, Early stage 12 wild-type (A,C,E), sna mutant (B,D) and esg;sna double mutant (F) embryos stained for Hnt and Crb (A,B), Hnt and dE-Cad (C,D) or Crb and Fas2 (E,F). (A,F) In both sna and sna,esg mutants endoderm cells undergo EMT; by stage 12 PMG cells are rounded and irregular in shape (F, compare with wild-type E), they have lost Crb expression (B,F, compare with wild-type A,C) and only express low levels of fragmented dE-Cad (D, compare with wild-type C). Scale bars, 20 μm.



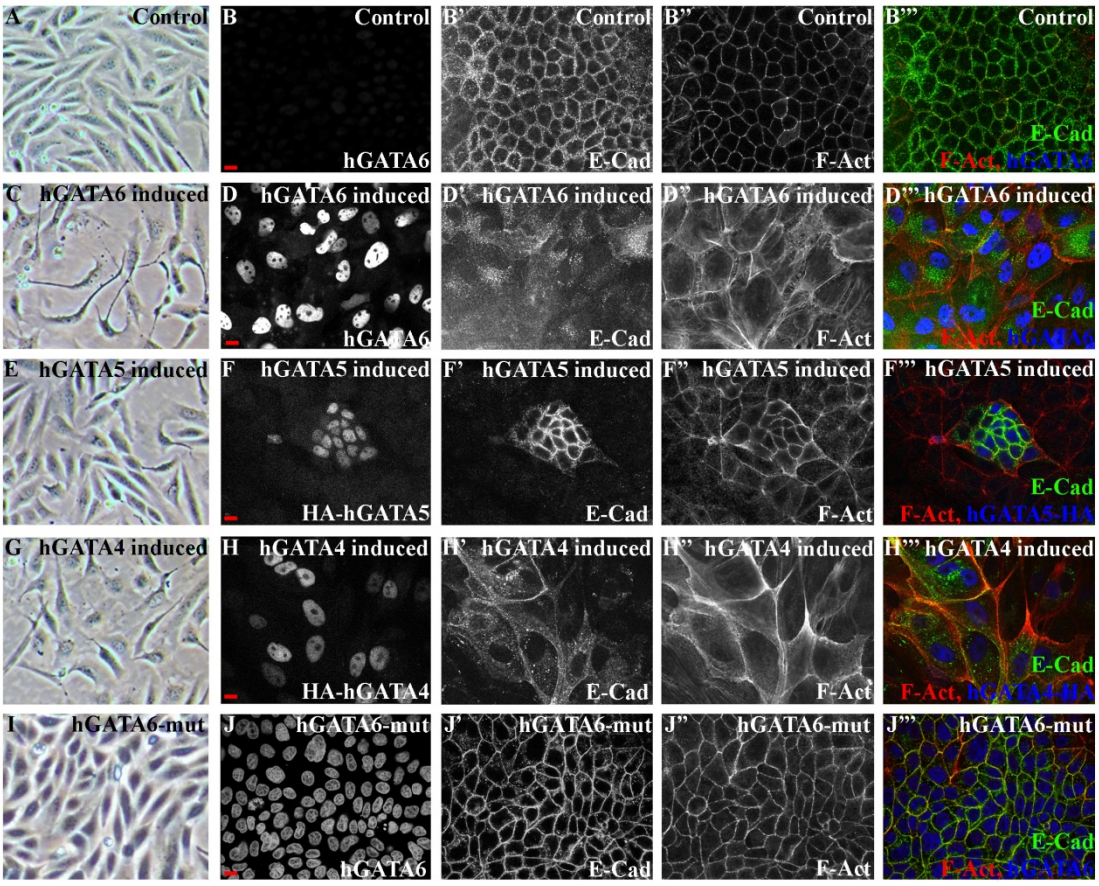

**Supplementary Figure S3. Ectopic expression of Srp in stripes of epidermal cells induces a loss of epithelial integrity and apical constriction.** A-G Stage 16 embryos expressing either GFP alone under control of the EngGal4 (A.C) or both Srp and GFP (B,D-G). Overexpression of Srp causes cells to lose their organisation, appear more dispersed and elongated in structure (compare B with A). (D,E) dE-Cad is delocalised around the cell cortex in Srp overexpressing cells found at the surface of the embryo (C,D, compare with wild-type C) F,G, a surface view (F) and a transverse confocal section (G) show that cells in contact with the outer epithelial layer are highly constricted at their apical surface (arrows). Scale bars, 20 μm.




**Supplementary Figure S4. Srp does not repress aPKC or dE-Cad transcription.** A-C *in situ* hybridization in wild type embryos for aPKC. aPKC remains strongly expressed in the PMG throughout stages 11 (A), 12 (B) and 13 (C). d,g, In stage 15 embryos the midgut forms a monolayered epithelial sheet enveloping the yolk (D) and dE-Cad is expressed throughout the PMG (G, arrow). e,h, Sna overexpression causes disruptions in midgut morphology (E, arrow) and a repression of dE-Cad (H, arrow). (F,I) Srp overexpression disrupts PMG migration so that the cells remain clustered at the posterior end of the yolk (F,I, arrow), but dE-Cad expression remains (I, arrow). J,K, In crb mutants dE-Cad becomes delocalised in PMG cells during stage 9 (compare K with wildtype J). Dotted lines outline the PMG.




**Supplementary Figure S5. Srp does not include EMT through the activation of Sna, Zfh1 or Zfh2.** A-D, Stage 16 embryos expressing both Srp and GFP under the control of EngGal4. A) Staining for Srp, confirms Srp is highly upregulated in stripes of epidermal cells. B-D) Ectopic expression of Srp does not induce expression of Sna (B), Zfh1 (C) or Zfh2 (D). E) Overexpression of Sna using EngGal4 has no effect on epidermal cells. Scale bar, 100 μm.




**Supplementary Figure S6. hGATA6 and hGATA4 induce EMT in MDCK cells, while hGATA5 does not.** A,B uninduced MDCK-hGATA cells. The cells are epithelial in morphology (A), do not express hGATA6 (B), and localise E-Cad and F-Actin to the apical junctions (B). c-h MDCK-hGATA6-dox (C,D), MDCK-hGATA5-dox (E,F) or MDCK-hGATA4-dox (G,H) cells 6 days after induction. (C,D) After addition of doxycycline hGATA6 protein accumulates in cell nuclei (D), the cells adopt a mesenchymal-like spindle shape morphology (C), the normal junctional localisation of E-Cad is lost (B') and F-Act appears reorganised into stress fibres and lamellipodia-like structures (D''). (E,F) After induction of HA-tagged hGATA5, which accumulates in cell nuclei (F) cells remain epithelial in morphology (E), and localise E-Cad and F-Act apically (F). (G,H) After induction of HA-tagged hGATA4, hGATA4 accumulates in cell nuclei (H), the cells adopt a mesenchymal-like spindle shape morphology (G), the normal junctional localisation of E-Cad is lost (H') and F-Act appears reorganised into stress fibres and lamellipodia-like structures (H''), similar to cells overexpressing hGATA6 (compare G,H with C,D).(I-J) MDCK cells transfected with a mutated form of hGATA6 accumulate hGATA6-mut in cell nuclei (J) but cells remain epithelial in morphology (I), and localise E-Cad and F-Act apically (J). Scale bars, 10 μm.

# SUPPLEMENTARY MOVIE LEGENDS

**Supplementary Movie S1.** Time-lapse movie of an embryo expressing UAS-GMA-GFP (from D. Kiehart) under the control of 48Y-Gal4. The movie begins at the start of germ band retraction, frames are taken every 90 s. (Quicktime, 7.4 MB). Data available on included DVD.

**Supplementary Movie S2.** Time-lapse movie of dE-Cad-GFP in endodermal cells as they undergo EMT and initiate migration. The movie begins around 5 hours of development, frames are taken every 60 s. (Quicktime, 2.9 MB). Data available on included DVD.

**Supplementary Movie S3.** Time-lapse movie of the wild type behaviour of stripes of epidermal cells. UAS-CD8-GFP is expressed under the control of Eng-Gal4, frames are taken every 90 s. (Quicktime, 2.4 MB). Data available on included DVD.

**Supplementary Movie S4.** Time-lapse movie of the behaviour of stripes of epidermal cells upon ectopic expression of Srp. Both UAS-Srp and UAS-CD8-GFP are expressed under the control of Eng-Gal4, frames are taken every 90 s. (Quicktime, 1.8 MB). Data available on included DVD.

**Supplementary Movie S5.** Time-lapse movie of the wild type behaviour of a subset of epithelial cells in the posterior of the embryo. UAS-GFP is expressed under the control of Cad-Gal4, frames are taken every 120 s. (Quicktime, 2.7 MB). Data available on included DVD.

**Supplementary Movie S6.** Time-lapse movie of the behaviour of a subset of epithelial cells in the posterior of the embryo upon ectopic expression of Srp. Both UAS-Srp and UAS-GFP are expressed under the control of Cad-Gal4, frames are taken every 120 s. (Quicktime, 1.6 MB). Data available on included DVD.

**Supplementary Movie S7.** Time-lapse movie of the wild type behaviour of a subset of epithelial cells in the anterior of the embryo. UAS-CD8-GFP is expressed under the control of Dll-Gal4, frames are taken every 120 s. (Quicktime, 6.2 MB). Data available on included DVD.

**Supplementary Movie S8.** Time-lapse movie of the behaviour of a subset of epithelial cells in the anterior of the embryo upon ectopic expression of Srp. Both UAS-Srp and UAS-CD8-GFP are expressed under the control of Dll-Gal4, frames are taken every 120 s. (Quicktime, 2.4 MB). Data available on included DVD.

# SUPPLEMENTARY EXPERIMENTAL PROCEDURES

## Fly strains and genetics

Unless otherwise noted, stocks were obtained from the Bloomington *Drosophila* Stock Center (Indiana University, USA). Wild type embryos were from *yw* stocks. *ubi-E-cad-GFP (II)* (from H. Oda) was used to follow dE-Cad in living embryos. The mutant strains used were $srp^{9L}$, $sna^{1}$ and $esg^{G66B}$, $sna^{1}$. For sustained/ectopic expression of Srp, UAS-Srp (from D. Hoshizaki) was expressed using the Gal4 drivers: 48Y-Gal4 (drives in the endoderm throughout embryonic development), Eng-Gal4 (drives in posterior compartments of the ectoderm), Cad-Gal4 (drives in the posterior of the embryo) or Dll-Gal4 (drives in the anterior of the embryo and some imaginal discs). For ectopic expression of Sna, UAS-Sna (from J. Kumar) was used. We maximised expression by collecting embryos at 29°C. Unless otherwise stated, when GFP is seen in confocal images, UAS-CD8-GFP was the co-expression marker. *crb*CRM-lacZ (from E- Furlong) was used to study Srp regulation of *crb* expression, and was previously described in (Sandmann et al., 2007).

## Immunohistochemistry, *in situ* hybridization, image acquisition and analysis

Antibodies used were: rabbit anti-aPKC (1:500; Hybridoma Bank); rabbit anti-Baz (1:1000; gift from A.Wodarz); rat anti-Crb (1:100; gift from E.Knust); rat DCAD2 (1:100; Hybridoma Bank); mouse anti-Dlg (1:500; Hybridoma Bank); mouse anti-Fas II (1:10; Hybridoma Bank); guinea pig anti-Fkh (1:250; ); goat anti-GFP (1:500; abCAM); mouse anti-Hnt (1:20; Hybridoma Bank); mouse anti-Mmp1 (a 1:1:1 of 3B8D12, 5H7B11 and 3A6B4 from the Hybridoma Bank was used at 3:10); Phalloidin TRITC (1:25; ); rabbit anti-SAS (1:500; gift from D.Cavener); rabbit anti-Sna (1:500; gift from M.Leptin); rabbit anti-Srp (1:1500; gift from D. Hoshizaki); rabbit anti-Zfh1 (1:750; gift from R.Lehmann); rat anti-Zfh2 (1:200; gift from C.Doe). For labelling with DCAD2 and Phalloidin embryos were fixed in 4% paraformaldehyde for 30 mins and devitillinised by hand. For all other stainings, embryos were fixed using standard techniques. Cy2, Cy3 and Cy5-conjugated secondary antibodies were from Molecular Probes and were used at 1:200 dilutions. *In situ* hybridization was performed according to standard protocols. The DaPKC cDNA used to generate the aPKC RNA probe was a gift from S. Sotillos (Sotillos et al., 2004). Probes were generated by DIG RNA labelling (Roche). Confocal images were acquired with a Leica SP5. Images were post-processed with Adobe Photoshop and ImageJ.

## Chromatin Immunoprecipitation (ChIP)

ChIP experiments were carried out as described previously (Sandmann et al., 2006). In each ChIP assay 4 independent chromatin samples were prepared from 0-6 hour wild-type embryo collections and 2 samples were incubated with anti-Srp sera and 2 samples without antibody, as a negative control. Each assay was preformed 3 times, and only regions that showed a strong enrichment in all 3 assays were considered to be positive. The following primers were used to assay for regions containing putative Srp-binding sites:

Hnt:
F-5'-AACGACAACAACTGGGACACCGC-3',
R-5'-TGCATCCGCAGAAAGGTGCGAA-3',
Crb1:
F-5'-ATTCTCGCCACTACCAACAGCGC-3',
R-5'- TAGTACAAGCTCAAGACGCGGCG-3';
Crb2:
F-5'-CATCCGGCTGTTGTTACGCGC-3',
R-5'-TGAGTTGGCCGCGCTTTTGCA-3';
Crb3:
F-5'-GAACCAACCTCTCTATAGGCCCTGT-3',
R-5'-GCCTTTCAGGGATTTCCCAGCCC-3';
Crb4:
F-5'-ACCATATTGTGTGGCCGAAGCGG-3',
R-5'-TGTTCCAACACCTCGTTGGCGAG-3';
Crb5:
F-5'-AACCGACCTGGCCTTGTCATGC-3',
R-5'-CGATGCACTGGAAGGAACGTGCA -3'.

## Quantification of RNA levels in the PMG

The PMG was dissected from wild-type embryos at 3-3.5 hours (Before), 6-6.5 hours (After) or from *srp* mutant embryos at 6-6.5 hours (*srp* mut), according to the method described in (Skaer, 1989). 3 PMGs were dissected per condition, RNA was isolated from the cells and amplified by Pico Profiling (Gonzalez-Roca et al., 2010). 4 separate sample sets were isolated, and the RNA amounts quantified by quantitative real-time PCR. The following SYBR green primer sets were used:
dE-Cad:
F-5'- CGGAGTGACGGGGCACGAAC-3',
R-5'-TGGTTCACACCGCCGAGGGA-3';
SAS:

F-5'-TTGGCCTCGCTGGGCGTATC-3',
R-5'-ACGCATCGTATTGCTGGTGCTGC-3';
Crb:
F-5'- CACGGAATGCTTGAACAACG-3',
R-5'-GTTTTGCTCACAGTGCTGAC;
Sdt:
F-5'- CAGTCCAGTAGTTCGGTTTGC-3',
R-5'-GATTTCAGCATCACCATTTCC-3';
Baz:
F-5'-AGATGAATAGATGGAGCAATCCC-3',
R-5'-GCATCGTCCTCGATAAGCAG-3';
aPKC:
F-5'- CATGATGACTGTGACTGTGG-3',
R-5'-CATTCCAATAAGAATTGCCCAG-3';
Par-6:
F-5'-CCACTCACGGCAGATGGGATGC-3',
R-5'-GGAGTGGTTGGGCGGCGATAC-3';
Hnt:
F-5'-ACCATCCGGCGGAGGATGCT-3',
R-5' -GCGGTGGCGTGGGACTGATC-3'

Reactions were carried out using SYBR Green Universal PCR Master Mix (Applied Biosystems) at an annealing temperature of 58°C. Reactions were carried out using a StepOnePlus Real-Time PCR System in clear 96-well reaction plates with optical covers, according to manufacturer's instructions. Gene expression levels were normalized using the endogenous control Actin5C for each sample and differences in target gene expression were determined using the StepOne 2.2 software.

**Cell Culture and Establishment of Cell Lines**

MDCK cells were stably transduced with lentiviral vectors containing the full length human Gata4, Gata5 and Gata6 with C-terminal HA-tags, and cloned downstream of the doxycycline inducible minimal CMV promoter of the pTRIPz vector backbone (Open Biosystems, RHS4743). A mutant full length Gata6 containing point mutation in its two zinc fingers (hGata6-mut) was also cloned in the same backbone and used as a negative control. Successfully infected cells were selected with 3 µg/mL puromycin and maintained under selection thereafter. Infection efficiency with the hGATA_pTRIPz vectors were over 50%.

## Western Blots

Cell lysates containing 20ug of protein were analyzed on either 10% or 7% SDS-PAGE gels to detect levels of E-cadherin and Actin, respectively. The antibody dilutions used were as follows: mouse anti-Ecadherin 1:1000 (BD, 610182) and rabbit anti-Actin 1:10000 (Sigma, A5060).

## Immunofluorescence of MDCK cells

MDCK cells were plated at a high confluence (150 000 cells/mL) and grown for 48h on sterile transparent 0.4µm pore polyester membrane inserts (Corning 3470) coated with 2.5 µg/cm$^2$ laminin (Sigma, L2020). Cells were fixed in 4% PFA for 10 minutes at room temperature, washed in PBS, incubated with 20mM glycine in PBS for 10 minutes, washed with PBS, permeabilized in 0.25% Triton-X 100 PBS for 20 minutes, washed with PBS and blocked with 1% BSA (Sigma) in PBS for 30 minutes. Samples were then incubated with the primary antibody for 1h at room temperature, washed in PBS, incubated with the secondary antibody for 1h at room temperature. Cells were then washed and nuclei were stained with 1µg/mL Falloidin-TRITC (Sigma, P1951), washed with PBS, and finally stained with DAPI (Sigma, D9542) at 0.1µg/mL for 10 minutes and washed with PBS. Slides were mounted in Vectashield Mounting medium (Vector). Antibodies used to label MDCK cells were: mouse anti-Ecadherin 1:100 (BD, 610182); mouse anti-Ncad 1:100 (BD 610920); goat anti-Gata6 1:200 (R&D, AF1700); rabbit anti-HA (Sigma, H6908). Confocal images were acquired using either a Leica TCS SPE or SP5 microscope. Secondary antibodies used were donkey anti-goat Dylight 488 and donkey anti-mouse Dylight 649; donkey anti-mouse FITC, Jackson). All images were acquired using a 63X magnification and image analysis was performed using LAS AF software (Leica). Images were processed and arranged in Adobe Photoshop and Adobe Illustrator.

## Invasion Assay

To obtain polarized monolayers for immunofluorescence, MDCK cells were seeded on transparent track-etched PET chambers with a surface area of 0.9 cm$^2$ (Falcon 353182) coated with 2-3mg/mL of matrigel (BD Biosciences, 353182) following manufacturer instructions. In brief, MDCK hGATA6 and control cells were seeded at a density of 5000 cells/well in DMEM medium supplemented with 0.5% BSA. Lower chambers were filled with 10% FBS to act as a chemoattractant. After 48 h of incubation, the cells on the lower membrane surface were fixed and stained with ice cold methanol for 5 minutes and stained with crystal violet (Sigma, HT90132). Eight randomly selected microscopic fields

(20X magnification) were imaged to calculate the average number of occupied pores per microscopic field. Results represent the mean plus/minus S.D. of at least two independent experiments performed in duplicate.

**Quantitative Real Time-PCR**

Total RNA samples were extracted using the TRIzol (Gibco BRL) and were DNAase treated, further purified using RNeasy columns (Qiagen) and reversed transcribed using the High Capacity cDNA Archive Kit (Applied Biosystems) according to the manufacturer's protocol. The following Canine specific SYBR green primer sets were used:

Gapdh:
F-5'-CATCACTGCCACCCAGAAG-3',
R-5'-CAGTGAGCTTCCCGTTCAG-3';
E-cad:
F-5'-AAGCGGCCTCTACAACTTCA-3',
R-5'-AACTGGGAAATGTGAGCACC-3';
Crb2:
F-5'- CTCCACCTCCCGTTCTTTC-3',
R-5'-CGAGACAAGTTCCACCATTG -3';
Crb3:
F-5'-ACGGCACCATTACACCCTCT-3',
R-5'-CGCTTCTCACGCAGTTTTC -3';
Pals1:
F-5'-TAGAACCCTTTACAGATGAGAG-3',
R-5'-ACTATCCGACTAATGATGACAG -3'
Patj:
F-5'-TGAGTTACCTGAAAGAGAAGAG-3',
R-5'-GTCCACCAACAATACTAATACC -3'
Cldn1:
F-5'-GATTCAGTGCAAGGTCTTCGACTCG-3',
R-5'- TGTGGCAACTAAAACAGCCAGACCT-3'
Cldn7:
F-5'-ATCCCGCAGTGGCAGATGAGCTCGT -3',
R-5'- CGGTCATAGCTATTCGGGCCTTCTT-3'
Cldn16:
F-5'-TGAAGCTGGTGGTAACTCGGGCATT-3',
R-5'- ACTGAACCAATGATTCCCGGGGCAC-3'
Mmp1:
F-5'-AAATCCCTTCTATCCGGAAGTTGAG -3',
R-5'- AAGCCAAAGGATCTGTGGATGTCCT-3'

Reactions were carried out using SYBR Green Universal PCR Master Mix (Applied Biosystems) at an annealing temperature of 62°C. Reactions were carried out using a StepOnePlus Real-Time PCR System in clear 96-well reaction plates with optical covers, according to manufacturer's instructions. Gene expression levels were normalized using the endogenous control GAPDH for each sample and differences in target gene expression were determined using the StepOne 2.2 software.

# CHAPTER 4

# HYPOXIA-DRIVEN SILENCING OF WNT TARGET GENES IN COLORECTAL CANCER STEM CELLS DURING DISEASE RELAPSE.

Gavin Whissell[1], Anna Merlos-Suarez[1*], Elisa Montagni[1*], Xavier Hernando-Momblona[1,] Evarist Planet[1], Marta Sevillano, Peter Jung[1], Alexandre Calon[1], Elena Sancho[1] and Eduard Batlle[1,2].

1. Oncology Program, Institute for Research in Biomedicine (IRB Barcelona), 08028 Barcelona, Spain
2. Institució Catalana de Recerca i Estudis Avançats (ICREA).

(*) These authors contributed equally to the work and their names are written in alphabetical order

Correspondence should be addressed to:
Eduard Batlle (eduard.batlle@irbbarcelona.org)

# SUMMARY

WNT signaling is required for the maintenance of colorectal cancer stem cells (CRC-SCs) but paradoxically low levels of WNT genes associate with a higher risk of disease relapse. Here we identify a core expression program driven by beta-catenin/TCF in CRC-SCs, which is a strict indicator of the dependency on WNT signaling for growth. We show that during disease progression part of this core program is downregulated as a consequence of tumor hypoxia. WNT/stem cell genes are prominently expressed in those CRC-SCs that form glandular structures whereas its silencing coincides with the upregulation of hypoxia-induced genes during tumor invasion. This transcriptional switch distinguishes those CRCs displaying highest probability of recurring upon therapy.

# INTRODUCTION

WNT signaling is essential for the maintenance of colon stem cells (CoSCs) (Schepers et al., 2012). In addition, the vast majority of colorectal cancers (CRCs) accumulate genetic alterations that activate the WNT pathway (Cancer Genome Atlas, 2012). As a consequence, CoSCs and CRCs share the expression of several beta-catenin/TCF target genes (van de Wetering et al., 2002a; Van der Flier et al., 2007). Despite mutational activation of the WNT pathway, the beta-catenin/TCF gene program is not expressed constitutively within CRCs. High levels of WNT genes such as LGR5 or EPHB2 define a subpopulation of cancer cells that display self-renewal and tumor-initiating capacity in mouse xenotransplantation assays (Kemper et al., 2012; Merlos-Suarez et al., 2011; Vermeulen et al., 2010). Furthermore, high levels of the gene expression signatures of CoSCs (Merlos-Suarez et al., 2011) or of CRC stem cells (CRC-SCs) (de Sousa et al., 2011) correlated with a higher risk of disease relapse upon therapeutic treatment in CRC patients. In sharp contrast with these data, it was shown that downregulation of CoSC/WNT target genes associated with poor outcome in the clinical setting (de Sousa et al., 2011). It has been proposed that CRC-SC abandon the CoSC-like phenotype to acquire an Embryonic Stem cell (ES)-like gene program during CRC progression (Kemper et al., 2012; Varnat et al., 2010). Apparently, this phenotypic change is mediated by methylation of WNT /CoSC target genes in aggressive tumors (de Sousa et al., 2011), a conclusion that was put forward from studying the behavior of small subset of WNT/CoSC genes.

Here we set out to test whether WNT/CoSC target genes are selectively methylated and assess the proposed interconversion from intestinal stem to ES-like state during CRC progression. Upon doing so, we have unraveled a strong link between WNT/CoSC gene regulation and hypoxia in aggressive CRC tumors. These findings identify a new player involved in the modulation of cancer stem cell phenotype.

# RESULTS AND DISCUSSION

To obtain a global view on how the WNT/CoSC program changes during tumor progression, we initially set out to broaden the catalogue of genes regulated by WNT signaling in CRC. We searched for WNT-driven genes in a collection of CRC cell lines derived from late stage tumors that carried activating mutations in the WNT signaling pathway. We used the strategy described in Figure 1A. In brief, each cell line was engineered to stably express the beta-catenin-binding domain of TCF4 (NTCF: amino acids 1 to 90) fused to a tamoxifen-inducible

version of the hormone-binding domain of the estrogen receptor (ERT2). Under basal conditions, this chimeric protein was retained in the cytoplasm (data not shown). Addition of 4-hydroxytamoxifen (4OHT) shuttled the NTCF-ERT2 fusion into the nucleus where it competed with endogenous TCFs for the binding to beta-catenin. This approach simulates the action of an ideal anti-WNT drug designed to disrupt the interaction of beta-catenin with TCFs. Activation of the NTCF-ERT2 fusion by 4OHT effectively blocked beta-catenin/TCF driven transcription in all cell lines as shown by decreased activity of the TCF reporter (TOP-FLASH) down to control (FOP-FLASH) values (red bars compared to purple bars in Figure 1B). We assessed the specificity of this approach through two controls; first, we generated the same CRC cells expressing ERT2. These cells did not display changes in beta-catenin/TCF reporter activity in the presence of 4OHT (Figure 1B). Second, we obtained Co115 CRC cells expressing either NTCF-ERT2 or ERT2. This cell line does not bear genetic alterations in the WNT signaling pathway and lacks beta-catenin/TCF transcriptional activity. We profiled changes in gene expression before and 36h after induction with 4OHT using microarrays. Induction of NTCF-ERT2 activity in Co115 cells did not modify the levels of any gene significantly (Figure S1C). In the rest of the cell lines the magnitude of transcriptional response induced by abrogation of beta-catenin/TCF transcriptional activity was largely heterogeneous and ranged from few to hundreds of probes significantly up- and down-regulated (Supplementary Figure S1C; page 126). 165 annotated genes (215 probes) were strongly downregulated (>2.5 fold, $p<0.05$) upon inhibition of beta-catenin/TCF activity in at least two cell lines of our collection (Figure 1C; Supplementary Table S1 (see DVD); see methods for details). We named this set the WNT ON program. Conversely, the WNT OFF program included those genes consistently upregulated upon inhibition of beta-catenin/TCF activity (435 annotated genes, 582 probes. Figure 1C; Supplementary Table S1 (see DVD)). Both WNT ON and WNT OFF gene sets remained unaltered in control ERT2-expressing CRC cells treated with 4OHT (data not shown).

Blockade of beta-catenin/TCF activity reduced the capacity of CRC cells to generate tumors in immunodeficient mice (Figure 1C-D and Supplementary Figure S1A; page 126). Yet, the degree of xenograft growth inhibition ranged from full suppression of tumorigenicity to marginal effects depending on the cell line (Figure 1C-D and Supplementary Figure S1A; page 126). We observed similar responses for *in vitro* proliferation (Supplementary Figure S1B; page 126). Importantly, average expression levels of the WNT ON program measured in basal conditions (i.e. prior to blockade of beta-catenin/TCF activity) correlated strictly with the dependency of each cell line on WNT signaling for growth (Figure 1D).

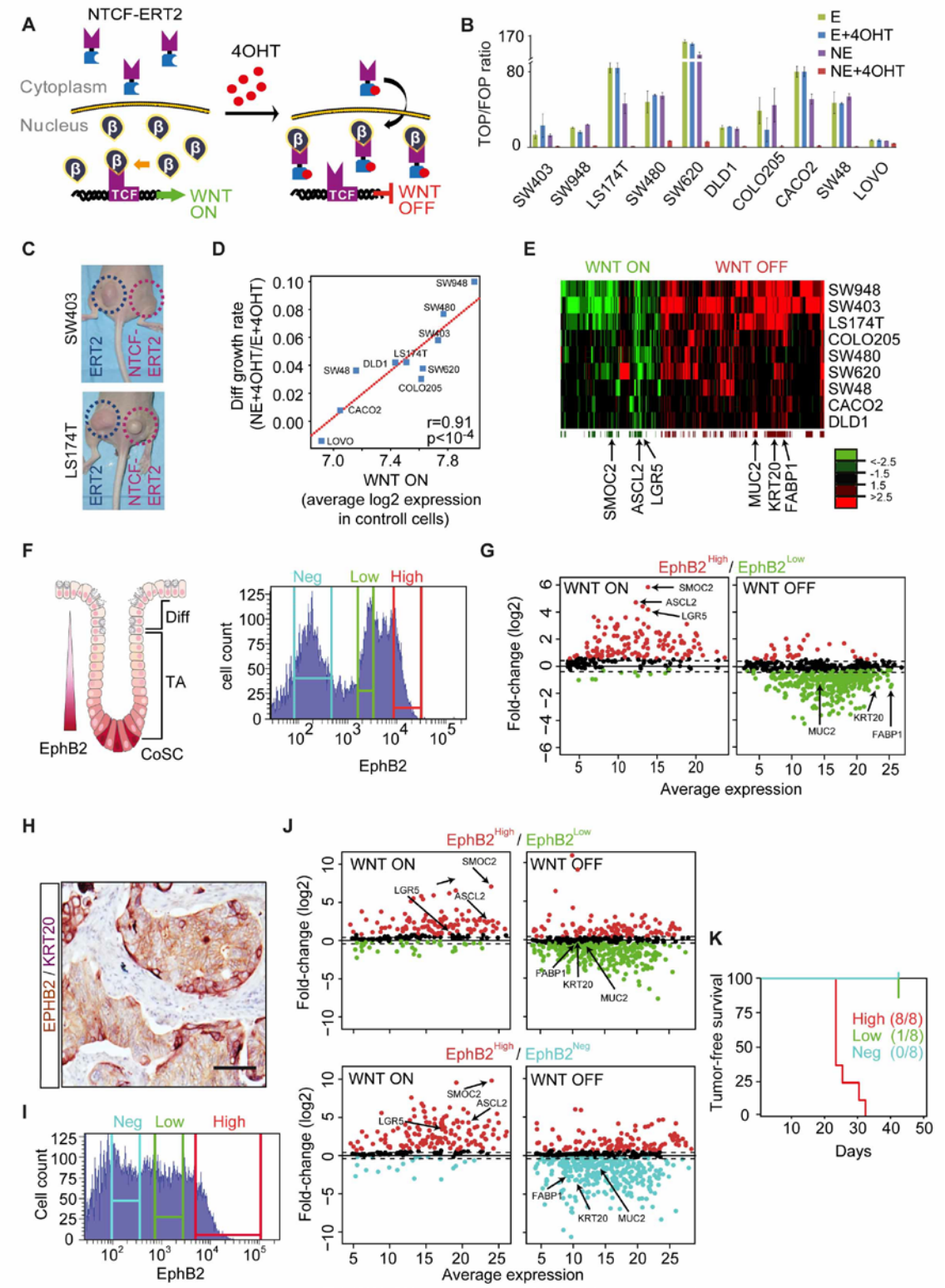



**Figure 1. Identification of the WNT target gene programs and characterization of their expression in CoSCs and CRC-SC. (A)** Schematic representation of the tamoxifen inducible NTCF-ERT2 fusion protein. (A) Beta-catenin/TCF reporter activity indicated by TOP-FLASH/FOP-FLASH reporter ratios (Morin et al., 1997) in CRC cell lines expressing ERT2 (E,

green bars), NTCF-ERT2 (NE, purple bars), ERT2 stimulated with 4OHT (E+4OHT, blue bars) or NTCF-ERT2 stimulated 4OHT (NE+4OHT, red bars). Reporter assays were performed in triplicate and representative assays are shown. Each bar represents the mean ± SD of 3 replicates. (B) Example of growth inhibition induced by blockade of beta-catenin/TCF activity in xenografts. TAM400 fed mice were injected with $1x10^6$ ERT2 or NTCF-ERT2 expressing cells in the left and right flanks, respectively. Mice were given 4OHT every 3 days, throughout the experiment. (C) Correlation between average expression of the WNT ON signature in control cells (ERT2) and reduction of xenograft growth rate upon inhibition of beta-catenin/TCF activity. The higher the expression of the WNT ON signature for a given cell line, the higher its dependency on WNT signaling for growth. Y axis shows decrease in tumor growth rate comparing ERT2- vs. NTCF-ERT2-expressing cells (fold change between the mean slope of ERT2 vs the mean slope of NTCF-ERT2, both treated with 4OHT) according to data shown in Supplementary Figure S1A; page 126. The red line is the least squares regression line. R indicates Pearson correlation estimate. (D) Cluster analysis showing relative changes in expression levels of the WNT ON and WNT-OFF programs in each NTCF-ERT2 vs ERT2 expressing CRC cell line 36h after addition of 4OHT. All the genes displaying fold changes >1.5 ($p<0.05$) in at least one cell line are represented. We also indicate well-characterized CoSC genes (SMOC2, LGR5 and ASCL2) and intestinal differentiation markers (MUC2, FABP1 and KRT20). (E) Left panel: Schematic representation of a human colon crypt displaying the EPHB2 gradient and the position of CoSC, TA cells and differentiated cells. Right: FACS profile according to EPHB2 surface levels of epithelial colonic cells (EPCAM+ gate) dissociated from healthy human colon mucosa. Red, green and blue areas indicate gates corresponding to EPHB2-high, -low and –negative cells. (F) Fold changes (Log2) in expression of each gene in the WNT ON and WNT OFF signatures between CoSCs (EPHB2-high) and Late TA cells (EPHB2-low) purified from normal human colon samples. Data shows the average expression obtained from comparing three independent normal human colon samples (Jung et al., 2011). (G) Double immunohistochemistry on a stage-IV primary CRC using anti-EPHB2 (brown) plus anti-KRT20 (magenta) antibodies. Scale bars = 10 µm. (H) FACS profile according to EPHB2 surface levels of epithelial colonic cells (EPCAM+ gate) dissociated from human stage IV CRC. Red, green and blue areas indicate gates corresponding to EPHB2-high, -low and –negative cells. (I) As in (G) but showing fold expression changes of WNT genes between EPHB2-high and -low or -negative tumor cell populations from CRC cells purified in (I). (J) Tumor free survival intervals of NOD/SCID mice (n=8) inoculated with 1000 EPHB2-high, -low or –negative cells isolated according to the FACS gates shown in (I). Frequencies of mice that developed tumors are indicated.

Amongst the WNT ON genes, we recognized several well-characterized CoSCs genes including *LGR5, SMOC2* or *ASCL2* (Figure 1E) (Barker et al., 2007; Munoz et al., 2012; van der Flier et al., 2009b). In addition, a substantial number of WNT ON genes have not been previously described as beta-catenin/TCF target genes. Prime examples are the hedgehog co-receptor PTCH1, the drug transporter ABCB1 or the immunomodulator CADM1, all of which play important roles during development and in cancer in other systems. The WNT OFF gene program included several differentiation markers such as *KRT20, FABP1* or *MUC2*. Fittingly, xenotransplanted CRC cells expressing the NTCF-ERT2 fusion construct in 4OHT treated mice formed tumor masses largely composed of cells displaying an intestinal differentiated phenotype (Supplementary Figure S1D; page 126). To further analyze the similarities between the WNT-driven program expressed in normal mucosa and in CRC, we took advantage of the technology recently developed by our laboratory that allows the purification of human CoSCs and their progeny from normal mucosa

samples (Jung et al., 2011). We sorted normal human colon mucosa epithelial cells (n=3) according to EPHB2 surface levels and obtained expression profiles of CoSCs (EPHB2-high), early Transient Amplifying (TA) cells (EPHB2-low) and Late TA cells (EphB2-neg) (Figure 1F) (Jung et al., 2011). This experiment indicated that the vast majority of the core WNT ON genes expressed by CRC cells were enriched in CoSCs whereas WNT OFF genes were upregulated in TA cells (Figure 1G). As in the case of the healthy mucosa, EPHB2 surface levels distinguishes tumor cells displaying different degrees of differentiation within human CRCs (Merlos-Suarez et al., 2011). CoSC-like tumor cells (EPHB2-high/KRT20-neg) and differentiated-like tumor cells (EPHB2-neg/KRT20+) occupied adjacent positions in CRCs (examples in Figure 1H and in Merlos-Suarez et al., 2011). We isolated these epithelial tumor cell populations (EPCAM+ and either EPHB2-high -low or -neg) by FACS from different human primary CRCs samples (Figure 1I and Supplementary Figure S1E-F; page 126) and profiled their expression. Comparative analysis revealed that most WNT ON genes were largely upregulated in EPHB2-high tumor cells (Figure 1J and Supplementary Figure S1E-F; page 126). Conversely, EPHB2-low and -neg tumor cells expressed highest levels of the WNT OFF program (Figure 1J and Supplementary Figure S1E-F; page 126). These results confirm that the WNT program is not homogenously expressed in CRCs. Paralleling the results with CRC cell lines, high expression of the WNT ON program identified those cell populations with greater tumorigenic potential, whereas adjacent tumor cells expressing the WNT OFF gene subset were benign as they did not propagate the disease upon xenotransplantation into immunodeficient mice (Figure 1K). Altogether, these observations demonstrate that CRC-SC retain a WNT program remarkably similar to that of normal CoSCs. We confirmed that activation of the WNT pathway is required to maintain the Stem/Undifferentiated phenotype as well as the tumorigenic potential of CRC cells (Scholer-Dahirel et al., 2011; van de Wetering et al., 2002a). Importantly, our analysis also revealed distinct degrees of addiction to beta-catenin/TCF signaling. This finding may be relevant to the design of upcoming therapeutic interventions using anti-WNT drugs. As we demonstrated for cell lines, CRC patients that may respond better to these therapies could possibly be identified using the WNT ON program described here.

Having identified the global set of genes regulated by beta-catenin/TCF in CRCs, we next investigated its modulation during disease progression. Medema and colleagues had showed that relative low expression levels of selected WNT ON/CoSC genes characterize a group of Stage II CRC patients with higher risk of undergoing disease relapse upon therapy (de Sousa et al., 2011). We extended this analysis to the full WNT ON program using the same patient cohort (Figure 2). The correlation of every gene of the WNT ON

signature with disease relapse is shown in Figure 2A and Supplementary Table S1 (see DVD). Indeed, a reduced yet significant subset of WNT ON genes (n= 19 genes out of 142. Green dots in Supplementary Figure 2A; page 128) inversely associated with disease relapse in this cohort (HR<-1, p<0.1) (Figure 2A). Yet, the same number of WNT ON genes (n= 19, Red dots in Figure 2A) were positively associated with relapse (HR>1, p<0.1 Figure 2A). As a result, the average expression of the full WNT ON program did not show a significant association with cancer recurrence in this cohort (HR for +1SD=-1.37, p=0.21). We made equivalent observations with a second independent CRC patient METAcohort that included 340 CRC cases at all disease stages (HR for +1SD =-1.16, p=0.26). We thus focused on the subset of WNT ON genes that negatively predicted relapse (HR<-1, p<0.1), from here named WNT $ON^{HR<-1}$ (Supplementary Table S1 (see DVD)). This included several previously reported CoSC-specific genes such as *APCDD1, AXIN2* or *ASCL2* (de Sousa et al., 2011). Decreased EPHB2 expression was also strongly associated with disease relapse (HR for +1SD=-2.16, p=0.02), confirming previous studies (Guo et al., 2006; Jubb et al., 2005; Lugli et al., 2005). In fact, *EPHB2* expression levels correlated directly with those of the WNT $ON^{HR<-1}$ genes (R=0.66, p<0.001; Figure 2B) implying that downregulation of EPHB2 occurs in more aggressive tumors concomitantly with the silencing of the WNT $ON^{HR<-1}$ genes (Figure 2B). The downregulation of WNT/CoSC genes probably reflects an important process during disease progression as we could observe this phenomenon in several independent CRC patient cohorts. Yet, it is paradoxical as the WNT ON/CoSC program confers tumorigenic potential to CRC cells and *EPHB2* is a marker gene for CRC-SC as shown in Figure 1 and elsewhere (Merlos-Suarez et al., 2011; Scholer-Dahirel et al., 2011; van de Wetering et al., 2002a; Vermeulen et al., 2010). As an explanation for this puzzling finding, it was proposed that CRC-SC switch from a CoSC-like phenotype characteristic of benign tumors to the expression of an Embryonic Stem cell (ES)-like gene program that is upregulated in aggressive neoplasms (de Sousa et al., 2011; Varnat et al., 2010). This hypothesis prompted us to re-analyze the data in Sousa et al. yet we could not demonstrate strong associations between the WNT and the ES cell programs in CRC. As a matter of fact, our analysis revealed that some ES cell-specific signatures (Ben-Porath et al., 2008) showed direct rather than inverse relationship with the WNT $ON^{HR<-1}$ levels in the CRC patient cohorts used in this study (Supplementary Table S2; page 130). On the contrary, genes repressed by the PRC2 complex in ES cells were inversely rather directly associated to WNT $ON^{HR<-1}$ levels in CRC (Supplementary Table S2; page 130). These trends were neither clearly associated to cancer recurrence in the both patient cohorts (Supplementary Table S2; page 130).

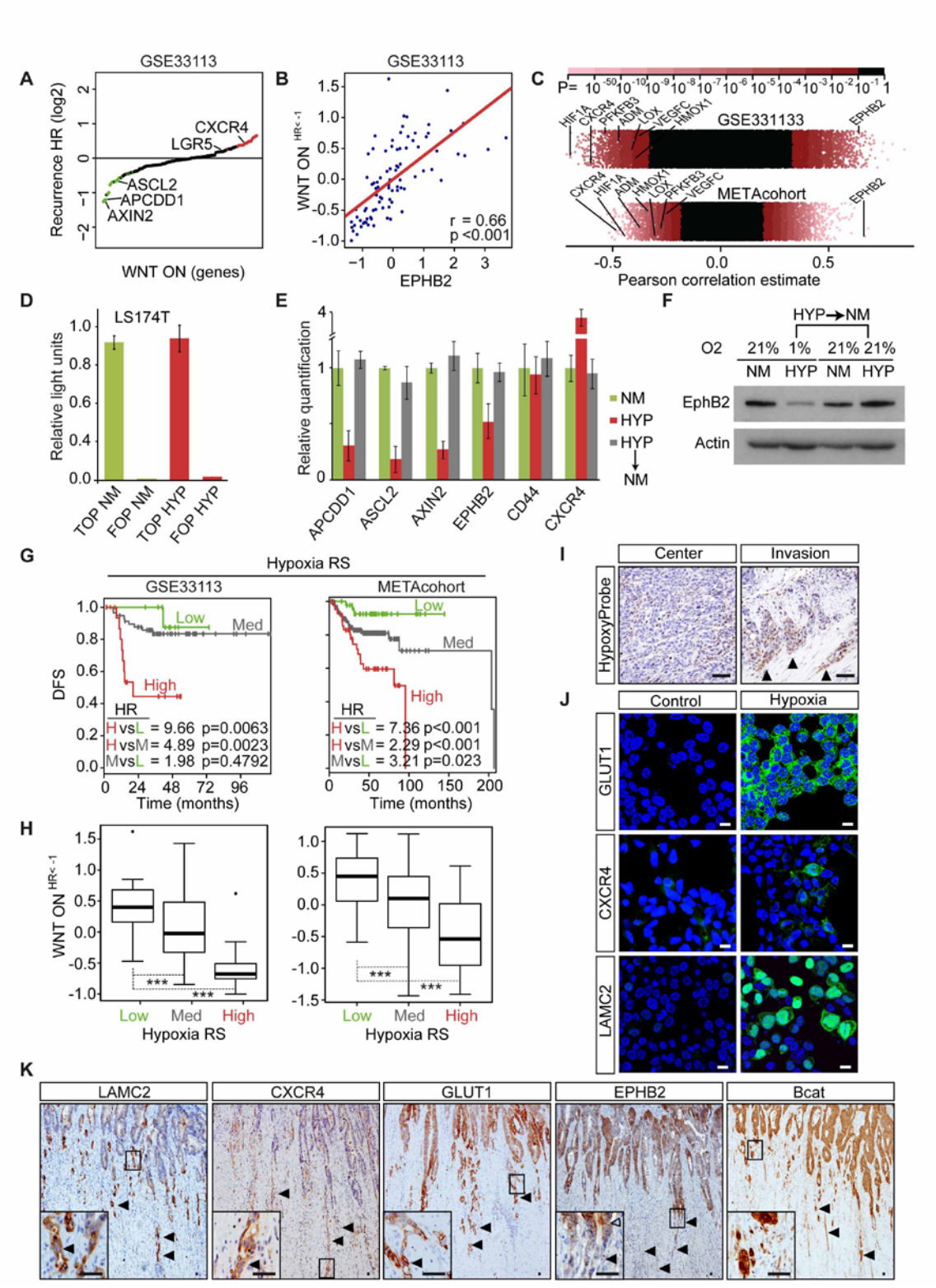



**Figure 2. Downregulation of a subset of WNT ON genes by hypoxia is associated with CRC relapse.** (A) Hazard Ratios (HR) for cancer relapse associated to each gene included in the WNT ON gene signature according to the GSE33113 CRC patient cohort. Green dots indicate WNT ON genes with HR<-1 and p<0.1, whereas red dots indicate genes with WNT ON genes with HR>+1 and p<0.1. (B) Correlation between *EPBH2* expression in each patient

(GSE33113 cohort) and average expression of the WNT ON$^{HR<-1}$ gene subset. Red line is the least squares regression line. R indicates Pearson correlation estimate. (C) Correlation of each gene present on the Affymetrix U133plus2.0 array platform to the average expression levels of the WNT ON$^{HR<-1}$ gene subset for both CRC patient cohorts. Color bar indicates p values in log scale. Genes displaying p value above 0.1 are colored in black. (D) Beta-catenin/TCF reporter activity (TOP-flash) and control reporter values (FOP-flash) of LS174T cultured under normal conditions (21% O2) (NM, green bars) or hypoxia (1%O2) (HYP, red bars) for 7 days. Reporter assays were performed in triplicate and representative assays are shown. Each bar represents the mean ± SD of 3 replicates. (E) Expression of key WNT ON genes in LS174T cells cultured under normal conditions (21% O2) (NM, green bars), hypoxia for 7 days (HYP red bars) or HYP for 7 days and then back to NM for 3 days (HYP to NM, grey bars). Values represent mean ± SD (n = 3). (F) Western-blot analysis using anti-EphB2 or anti-actin antibodies on lysates from LS174T cells cultured for 7 days in 21% O2 (NM) or 1% O2 (HYP) or HYP for 7 days and then back to NM for 9 days. (G) Recurrence free survival analyses for patients in the GSE33113 or METAcohort that express different average levels of the HYP-RS. Low levels correspond to expression below -1SD (green line), medium for patients with expression between -1SD and +1SD (grey line) and high with expression higher than +1SD. (H) Boxplot of the expression of the WNT ON$^{HR<-1}$ gene subset for patients in the GSE33113 or METAcohort with low, medium or high expression of HYP-RS as described in (G). (I) Immunohistochemistry using Hypoxyprobe on tumor center and invasion area of a representative orthotopic murine caecum tumor. Scale bars = 50 μm (J) Immunofluorescence of three marker genes of hypoxia (GLUT1, CXCR4 and LAMC2) in LS174T cells cultured in 21% O2 (NM) or 1% O2 (HYP) for 7 days. Scale bars = 10 μm. (K) Immunohistochemistry of a representative human CRC invasion front for LAMC2, CXCR4, GLUT1, EPHB2 and beta-catenin (Bcat). Scale bars = 25 μm. A total of 34 patient samples with GLUT1 positive invasion fronts were scored for LAMC2 (91%) and CXCR4 (75%) positivity.

As part of the transition from a CoSC-like to an ES-like phenotype during CRC progression, it was proposed that WNT ON$^{HR<-1}$ genes were silenced by methylation (de Sousa et al., 2011). This conclusion was reached upon assessing methylation of a limited subset of CpG islands in the promoter of selected WNT/CoSCs genes (de Sousa et al., 2011). To broaden this analysis, we took advantage of the global methylation analysis performed as part of Cancer Genome Atlas project (Cancer Genome Atlas, 2012). This study examined 222 CRC samples using the Infinium Methylation platform (Ilumina, Inc) that measures methylation of several CpG sites located in the proximity of the transcription start sites of 14,475 consensus coding sequences (Bibikova et al., 2009). Mining of this dataset revealed that 42 % of assessed WNT ON$^{HR<-1}$ genes displayed significant levels of methylation in CRCs (Supplementary Figure S2A-left; page 128 and Supplementary Table S1 (see DVD)). *EPHB2* was not significantly methylated (data not shown). Next, we split the WNT ON$^{HR<-1}$ genes into methylated and not methylated and compared the predictive power of these two subsets for disease progression. In both CRC patient cohorts, reduced expression levels of non-methylated WNT ON$^{HR<-1}$ subset had equivalent or better predictive power than that of the methylated subset (Supplementary Figure S2B; page 128). We also investigated the methylation status of the WNT ON genes that showed positive association with cancer recurrence (i.e. those represented as red dots in Figure 2A, WNT ON$^{HR>+1}$).

More than half of the genes in this group were also highly methylated (Supplementary Figure S2A-right; page 128), despite the fact that they predicted poor prognosis. Importantly, their degree of methylation correlated directly to that of the WNT ON$^{HR<-1}$ subset in CRC (R=0.31, p <0.001. Supplementary Figure S2C; page 128). Our analyses also indicated that the ratio of methylation of both the WNT ON$^{HR<-1}$ and WNT ON$^{HR>+1}$ subsets in each patient followed genome-wide methylation rates (Supplementary Figure S2D; page 128). These data argue against the existence of a mechanism dedicated to specifically methylate WNT genes in CRC. We conclude that whereas a fraction of WNT target genes becomes frequently methylated in those CRCs that display high rates of global methylation (CIMP-high), this phenomenon occurs independently of their association with disease relapse.

To identify the mechanism responsible for the downregulation of the WNT/CoSCs program during disease progression, we searched in an unbiased manner for genes and pathways prominently upregulated in CRCs displaying low WNT ON$^{HR<-1}$ levels. We discovered that several well established hypoxia-induced genes such as *LOX* (Erler et al., 2006), *VEGFC* (Chaudary and Hill, 2009), *ADM* (Frede et al., 2005), *HMOX1* (Lee et al., 1997), *PFKFB3* (Minchenko et al., 2002) or *CXCR4* (Staller et al., 2003) as well as the hypoxia-signaling mediator *HIF1A* were inversely and very significantly correlated to WNT ON$^{HR<-1}$ average expression levels in the two CRC patients cohorts used in this study (Figure 2C). This observation prompted us to test the effect of hypoxia on WNT target gene expression. To this end, we cultured LS174T CRC cells in low oxygen concentration (1%). Hypoxia did not modify beta-catenin/TCF transcription as measured by reporter assays (Figure 2D). Yet, we observed decreased levels of several WNT ON genes including most of the WNT ON$^{HR<-1}$ subset (Figure 2E). Of note, even those WNT ON$^{HR<-1}$ genes displaying the highest methylation rates in CRC (*ASCL2, APCDD1* and *PTPRO*) were downregulated in hypoxic conditions (Figure 2E). Hypoxia also controlled *EPHB2* mRNA (Figure 2E) and protein levels (Figure 2F), as previously suspected (Kaidi et al., 2007). Interestingly, this expression switch was reversible as WNT ON$^{HR<-1}$ genes returned to original levels upon re-exposure of cells to 21% oxygen (Figure 2E). To further explore the relationship between hypoxia, WNT gene expression and disease outcome in CRC, we identified the set of genes upregulated by LS174T CRC cells during hypoxia (>2 fold, p<0.05. Supplementary Table S1 (see DVD)). This hypoxia-response signature (HYP-RS) included well-established hypoxia markers such as PFKFB3 and SLC2A1 (also known as GLUT1). To avoid biases in subsequent analyses, we removed all WNT ON and WNT OFF genes from the HYP-RS. We found that elevated average levels of the HYP-RS were remarkably robust predictors of disease relapse in the two CRC patient cohorts (Figure 2G). A second HYP-RS

generated by equivalent means from tumor organoids expanded *in vitro* from primary CRC-SC (Supplementary Table S1 (see DVD)) also predicted disease progression (Supplementary Table S2; page 130). Importantly, HYP-RS-high tumors displayed relatively low WNT $ON^{HR<-1}$ average levels (Figure 2H). They also upregulated several key mediators of tumor cell invasion such as ZEB2 and SNAI2, as well as secreted factors that promote tumor cell motility including HGF and SDF1 (Supplementary Figure S2E; page 128). Therefore, low O2 concentration in CRC is coupled to a more aggressive phenotype that includes reduced WNT target gene expression.

To conclude, we tracked the localization of hypoxic cells within tumors. Mice carrying distinct orthotopic CRC xenografts were administered the hypoxia probe pimonidazole hydrochloride prior to sacrifice. Detection of Hypoxyprobe on tissue sections demonstrated higher accumulation at invasive fronts compared to the central area of the tumor in all xenografts (example in Figure 2I). Consistent, with this observation, we found that several hypoxia-induced genes in CRC cells including *LAMC2, CXCR4* and *GLUT1* (Figure 2J) were more prominently expressed at invasive fronts of human CRC samples (Figure 2K). In contrast, EPHB2 expression was restricted to clusters of tumor cells located at regions displaying a cohesive and epithelial organization. They were predominantly located in tumor glands that occupied the central areas of most CRCs. Yet, EPHB2 did not mark cells that separated from the tumor bulk and that invaded the adjacent stroma (Figure 2K and Supplementary Figure S2G; page 128) as previously reported (Guo et al., 2006; Karamitopoulou et al., 2010). Most of these invasive tumor cells displayed obvious accumulation of beta-catenin in the nucleus (Figure 2K) (Brabletz et al., 2001) and expressed low levels of the epithelial maker EPCAM (Supplementary Figure S2F; page 128). Similar to the central regions of the tumor (Merlos-Suarez et al., 2011; Ziskin et al., 2012), tumor buds contained either KRT20(+) or KRT20(-) cells likely owing to the heterogeneous differentiation states of invasive CRC cells (Supplementary Figure S2H-I; page 128).

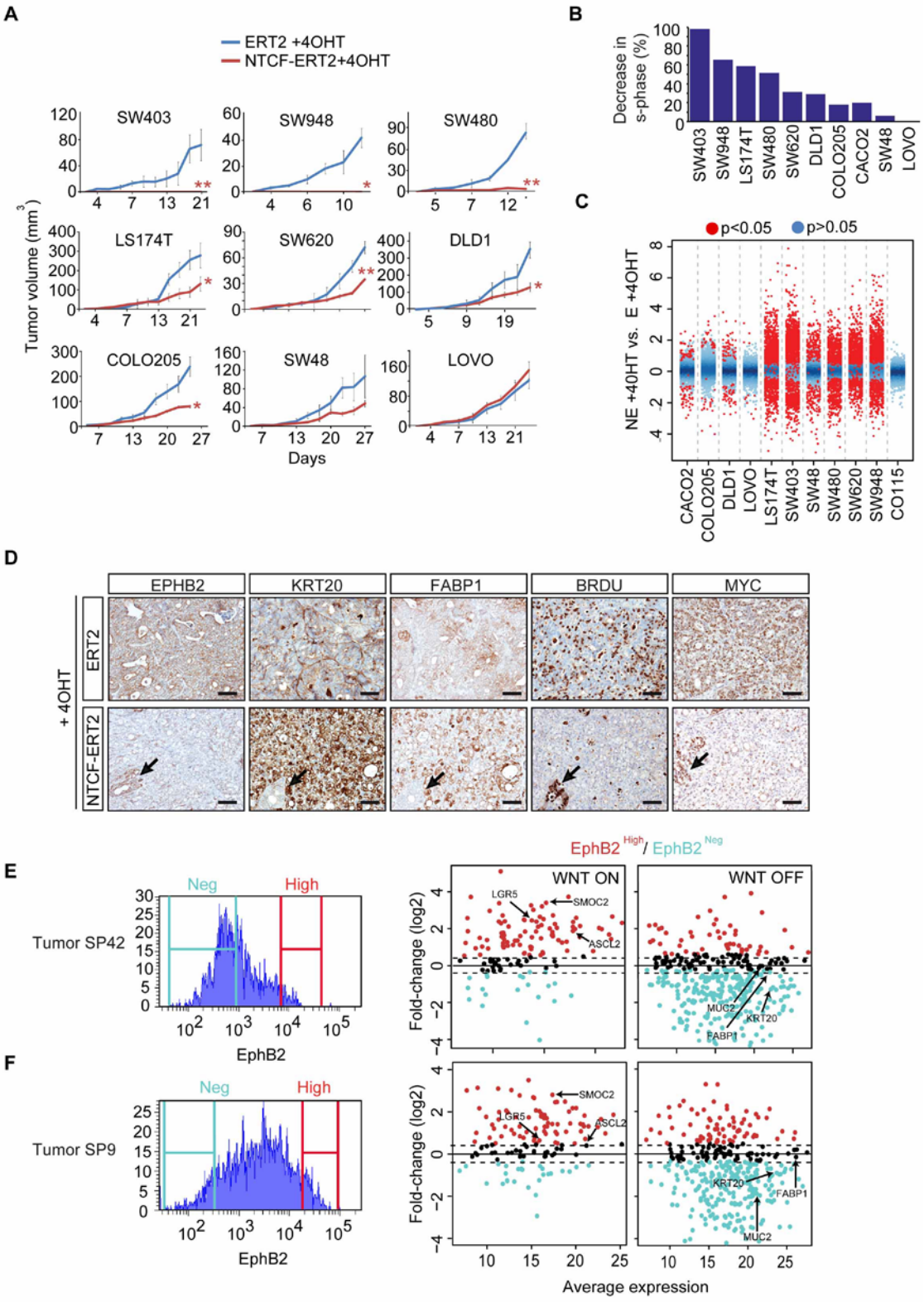
Our and other laboratories had demonstrated that EPHB/ephrin-B signaling blocks tumor progression by limiting cell invasion and enforcing E-cadherin-mediated adhesion both *in vitro* and *in vivo* (Batlle et al., 2005; Chiu et al., 2009; Cortina et al., 2007). Here we show that hypoxia decreases EPHB2 as well as levels of several WNT/CoSC genes, a phenomenon that robustly predicts disease relapse upon therapy. Hypoxia-induced genes are expressed at highest levels at invasion fronts, coinciding with silencing of *EPHB2*. The reversible nature of hypoxia-mediated downregulation of WNT/CoSC genes fit in well with the notion proposed by Brabletz and colleagues that CRC metastases reacquire a cohesive epithelial-like organization that resembles that of the primary tumor

(Brabletz et al., 2001; Brabletz et al., 2005). Altogether these observations provide a rational for the unexpected finding that poor prognosis in CRC associates with decreased levels of WNT/CoSC genes and suggests a transition from static to migratory phenotype induced by hypoxia in CRC-SCs.

## ACKNOWLEDGMENTS

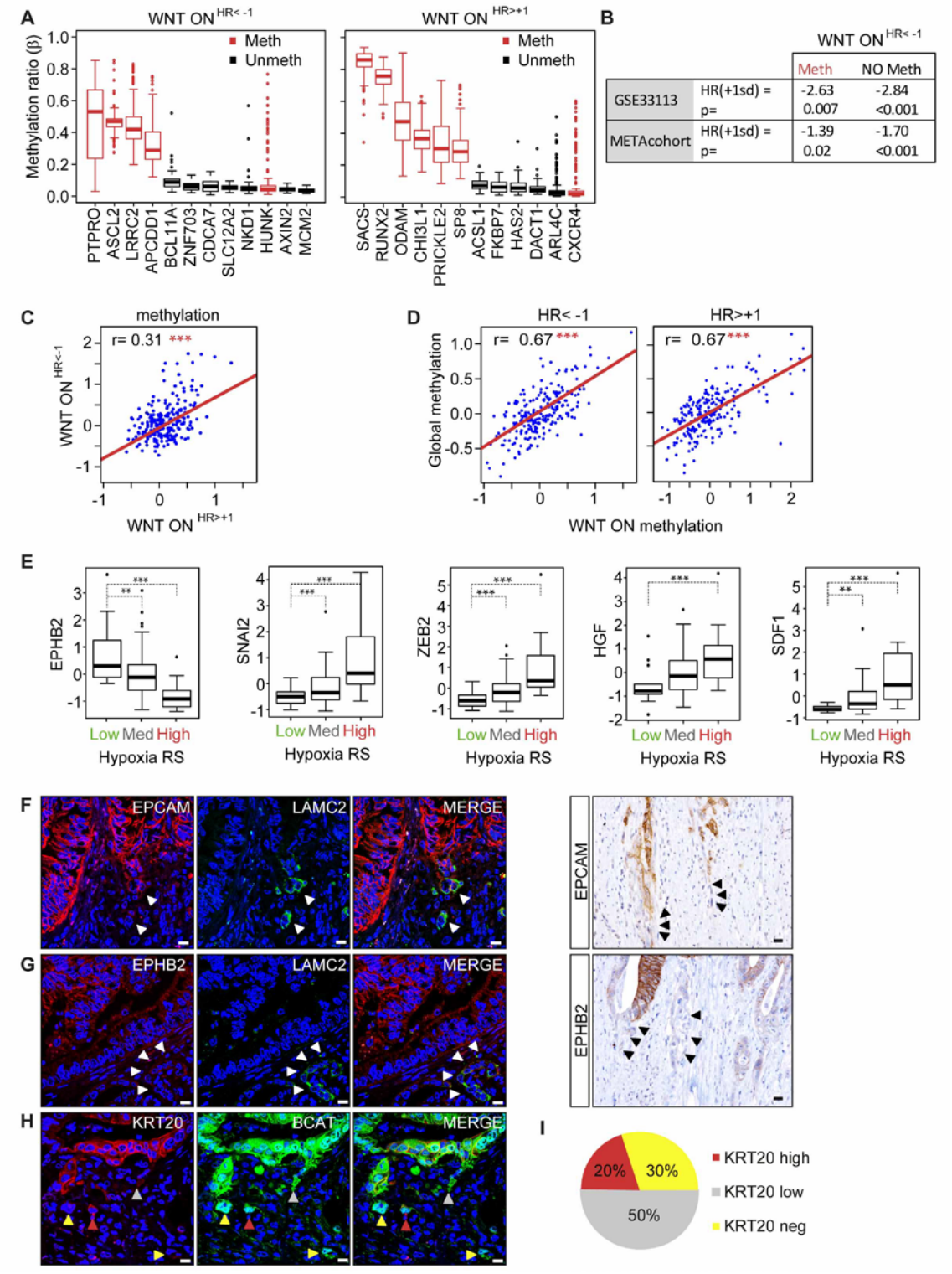
We thank the IRB Functional Genomics Core Facility for technical assistance in microarray hybridization experiments, David Rossell from the Biostatistics and Bioinformatics Unit for support with data analysis and Daniele Tauriello and Sergio Palomo Ponce for providing orthotopic tumor samples and the rest of the Batlle lab for discussions. This work was supported by grants to E.B. from the European Research Council (Starting grant - 208488) and MICINN (Plan Nacional I+D+I and Consolider programmes) and to A.M-S. from Fundación Olga Torres.

# SUPPLEMENTARY FIGURES




**Supplementary Figure S1. Effects on beta-catenin/TCF blockade on tumor growth and morphology.** (A) Growth of subcutaneous xenografts generated by NTCF-ERT2 (red) or ERT2

expressing cells (blue) in the flanks of immunodeficient mice that were fed TAM400 and injected i.p. with 4OHT every 3 days during the course of the experiment. Stars indicate statistically significant differences of the log growth rate between ET and NET using a t-test (* $p<0.05$ and ** $p<0.01$). (B) Reduction of cells in S phase upon blockade of beta-catenin/TCF *in vitro*. Graph shows the percent decrease in NTCF-ERT2-expressing cells compared to ERT2-expressing cells cultured with 4OHT for 36 hours. Representative values for each cell line are shown. (C) Representation of the changes in gene expression induced by inhibition of beta-catenin/TCF activity in CRC cell lines. Probes that changed at least 1.5 fold in one out of 11 cell line were selected. Blue dots indicate probes displaying non-significant changes ($p>0.05$) in NTCF-ERT2 compared to ERT2-expressing cells both cultured with 4OHT for 36 hours whereas red dots indicate probes showing significant changes ($p<0.05$). Note no significant changes in control CRC cell line CO115 lacking BCAT/TCF activity. P-values were obtained through moderated t-tests as implemented in the limma package of bioconductor. P-values were adjusted for multiple testing with the Benjamini and Hochberg method. (D) Analysis by immunohistochemistry of marker gene levels on serial sections of subcutaneous tumors generated by SW403 CRC cells five days after IP injection with 4OHT. We show stem /progenitor cell marker (EPHB2), intestinal differentiation markers (FABP1 and KRT20) or proliferation markers (BRDU and MYC). Arrows indicate small cluster of cells that escaped from beta-catenin/TCF inhibition. Scale bars = 50 µm. (E) Example of stage II CRC sample. EPHB2 surface expression levels of tumor epithelial cells (EPCAM+ gate). Log2 fold-change of the core WNT_ON and WNT_OFF in EPHB2-high vs. EPHB2-negative cells purified from the abovementioned CRC. (F) Example of stage II CRC sample. EPHB2 surface expression levels of tumor epithelial cells (EPCAM+ gate) demonstrate a majority of EPHB2-negative cells. Log2 fold-change of the core WNT ON and WNT OFF in EPHB2-high vs. EPHB2-negative cells purified from the abovementioned CRC.




**Supplementary Figure S2 – Hypoxia drives the silencing of a subset of WNT target genes irrespective of methylation.** (A) Methylation ratios (β-values) in the Infinium Methylation platform in the GCAT cohort of WNT ON$^{HR<-1}$ (left) or WNT ON$^{HR>+1}$ (right) genes. The β-values is a ratio between Illumina methylated probe intensity and total probe intensities (sum of methylated and unmethylated probe intensities). It is in the range of 0 and 1, which can be interpreted as the percentage of methylation. We considered a gene is methylated when 5% of

its samples have methylation values above 0.25 (red boxes). This threshold is arbitrary but it represents a relaxed estimation of those genes with a significant degree of methylation. (B) WNT_ON$^{HR<-1}$ were divided into methylated (Meth) and not-methylated (NO Meth) according to the criteria indicated in (A) and the average expression of each subset was used to predict cancer recurrence in GSE33113 and the METAcohort. The predictive value for risk of disease relapse shown as HR for every increase in standard deviation in expression. (C) Correlation between average methylation score of WNT_ON$^{HR<-1}$ (A-left, red boxes) and WNT_ON$^{HR>+1}$ genes (A-right, red boxes) analyzed in the Infinium Methylation platform in the GCAT patient cohort (n=222 samples)(Cancer Genome Atlas, 2012). Each dot represents a tumor sample. Only methylated genes according to the criteria used in panels A were used for the analysis. r indicates Pearson correlation estimates. The red line is the least squares regression line. *** pvalue <0.001. (D) Correlation between average methylation score of WNT_ON$^{HR<-1}$ (A-left, red boxes) or WNT_ON$^{HR>+1}$ genes (A-right, red boxes) and the average methylation rates of the rest of genes analyzed in the Infinium Methylation platform in the GCAT patient cohort (n=222 samples) (Cancer Genome Atlas, 2012). Each dot represents a tumor sample. Only methylated genes according to the criteria used in panels A were used for the analysis. r indicates Pearson correlation estimates. *** pvalue <0.001. (E) Expression levels of *EPHB2* in GSE33113 and genes involved in tumor cell invasion (*SNAI2* and *ZEB2*) and secreted factors that promote tumor cell motility (*HGF* and *SDF1*) in patients with low, medium or high average levels of the HYP-RS as defined in Figure 1G. (F) Example of lack of EPCAM staining at invasion fronts. Pictures show double immunofluorescence for the epithelial maker EPCAM (red) and the marker of invasive cells LAMC2 in an invasion front of a representative CRC. Note that tumor buds are LAMC2+ but express low or no levels EPCAM. This pattern was consistently observed in >100 invasion fronts corresponding to 34 different tumor samples. Immunohistochemical analysis of EPCAM expression in a representative CRC invasion front (right panel). Notice EPCAM low/negative invading cells budding into the stroma (black and white arrowheads). Scale bars = 10 μm. (G) Example of loss of EPHB2 expression at invasion fronts. Pictures show double immunofluorescence for EPHB2 (red) and the marker of invading cells LAMC2 (green) in an invasion front of a representative CRC. Note that EPHB2 decorates tumor glands yet is absent from tumor buds that stain for LAMC2. This pattern was consistently observed in >100 invasion fronts corresponding to 34 different tumor samples. Immunohistochemical analysis of EPHB2 expression in a representative CRC invasion front (right panel). Notice EPHB2 invading cells budding into the stroma (black and white arrowheads). Scale bars = 10 μm. (H-I) The pan-differentiation marker KRT20 (red) is heterogeneously expressed in tumor buds at invasive fronts. Cancer cells were recognized by nuclear beta-catenin accumulation (green). (H) We identified cells within tumor buds that express high (red arrow), low (yellow arrowhead) or negative KRT20 levels (purple arrowhead). Scale bars = 10 μm. (I) Percentage of cells in tumor buds that express high, low or negative KRT20 levels identified in >100 invasion fronts corresponding to 34 independent CRC samples. Red, purple and yellow arrowheads point to nuclear beta-catenin positive KRT20 high, low and negative cells, respectively.

# SUPPLEMENTARY TABLES

**Supplementary Table S1. Gene signatures.** WNT ON, WNT OFF, HYP-RS (LS174T), HYP-RS (primary tumor organoid), WNT ON HR<-1 and WNT ON HR>+1. Data available on included DVD.

**Supplementary Table S2. Predictive power of ES-cell specific gene expression signatures and correlation with the WNT_ON$^{HR<-1}$ gene subset in CRC.** We analyzed the predictive value for risk of disease relapse shown as HR per every increase in standard deviation in expression and the correlation with average expression of WNT_ON$^{HR<-1}$ subset of each gene expression signatures specific of ES cells in two CRC patient cohorts. These ES-derived signatures had been previously described and predict disease progression in breast cancer (Ben-Porath et al., 2008). They correspond to: (i) ES expressed genes (ES exp1 and exp2): two sets of genes overexpressed in ES cells compared to other cells and tissues according to a multistudy compilation and meta-analysis (Assou et al., 2007).(ii) NANOG, OCT4 and SOX2 targets: four sets of genes whose promoters are bound and activated in human ES cells by each of these regulators of ES cell identity, or co-activated by all three (Boyer et al., 2005), and an additional set (NOS TFs) including a subset of NOS activation targets encoding transcription regulators; (iii) Polycomb targets (SUZ12, Eed or PRC2 targets): four sets representing genes bound by the Polycomb repressive complex 2 (PRC2) in human ES cells (Lee et al., 2006). (iv) MYC targets: two sets of genes bound and activated by MYC, identified in two independent studies (Fernandez et al., 2003). (v) For comparison, we also include the Hypoxia Response Signature in LS174T and CRC-SC-derived tumor spheres. According to the hypothesis of a CoSC-like to ES-like phenotypic transition in aggressive CRCs (de Sousa et al., 2011) we would have expected that upregulation of ES cell specific gene expression signatures (ES_exp1, ES-exp2), NANOG-OCT4-SOX2 target genes and MYC targets would have been associated with cancer recurrence and inversely correlated to the WNT_ON$^{HR<-1}$ subset. On the contrary, polycomb targets should have been inversely associated with relapse as well as directly to WNT_ON$^{HR<-1}$ subset as PRC2 represses differentiation genes in ES cells. Yet, we did not find any consistent pattern in the two patients cohorts supporting that ES-cell programs are general predictors of disease progression in CRC or that CoSC-like tumors silence WNT/CoSC genes to acquire ES cell traits. We have highlighted analyses (P<0.05) that predict poor and good prognosis in red and green, respectively.

| | METAcohort | | | | | GSE33113 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | HR for relapse (+1SD) | p value | Spearman correlation with WNT ON $^{HR<-1}$ | p value | | HR for relapse (+1SD) | p value | Spearman correlation with WNT ON $^{HR<-1}$ | p value |
| **ES exp1** | -1.28 | 0.0575 | **0.182** | **0.0022** | | -1.85 | 0.014 | 0.139 | 0.1946 |
| **ES exp2** | -1.35 | 0.0214 | **0.382** | **<0.0001** | | -3.32 | <0.001 | **0.525** | **<0.0001** |
| **NANOG targets** | -1.01 | 0.9399 | **0.157** | **0.0083** | | 1.12 | 0.6321 | -0.043 | 0.6879 |
| **OCT4 targets** | 1.68 | 0.0009 | -0.095 | 0.1104 | | 1.65 | 0.0368 | **-0.326** | **0.0019** |
| **SOX2 targets** | 1.11 | 0.4596 | **0.176** | **0.0031** | | 1.15 | 0.5443 | -0.069 | 0.5196 |
| **NOS targets** | 1.62 | 0.0015 | **-0.234** | **<0.0001** | | 1.43 | 0.0998 | **-0.289** | **0.0062** |
| **SUZ12 targets** | 1.13 | 0.398 | **-0.267** | **<0.0001** | | 1.44 | 0.0844 | **-0.318** | **0.0025** |
| **EED taregts** | 1.12 | 0.4192 | **-0.214** | **0.0003** | | 1.34 | 0.163 | **-0.372** | **0.0003** |
| **PRC2 targets** | 1.01 | 0.9565 | **-0.212** | **0.0004** | | 1.35 | 0.1649 | **-0.323** | **0.0021** |
| **MYC1 targets** | -1.2 | 0.1556 | **0.177** | **0.0029** | | -1.33 | 0.2354 | 0.148 | 0.1661 |
| **MYC2 targets** | -1.24 | 0.0895 | **0.198** | **0.0009** | | -1.36 | 0.2022 | 0.144 | 0.1793 |
| | | | | | | | | | |
| **Hypoxia-RS (LS174T)** | 1.97 | <0.0001 | **-0.475** | **<0.0001** | | 1.71 | 0.0182 | **-0.57** | **<0.0001** |
| **Hypoxia-RS (primary CRC-SCs tumor spheroids)** | 1.88 | <0.0001 | **-0.433** | **<0.0001** | | 1.71 | 0.0159 | **-0.621** | **<0.0001** |
| | | | | | | | | | |

# SUPPLEMENTARY EXPERIMENTAL PROCEDURES

## Cell culture

CACO2, COLO205, CO115, DLD1, LS174T, LOVO, SW403, SW48, SW480, SW620 and SW948 were purchased from the American Type Culture Collection (ATCC, USA). All cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with L-Glutamine (Invitrogen 41966, Paisley, Scotland) and 10% FBS (Invitrogen 10270) at 37ºC and 5% CO2. For hypoxia studies, cells were grown in DMEM media containing 1% FBS. Hypoxic conditions (HYP) (1% O2) were achieved using a hypoxia incubator (Thermo Electron, Forma, MA) set to 37ºC and possessing 5% CO2. Cells were grown for 7 consecutive days. Control cells were grown under serum starved conditions as above and placed under normal (NM) (21% O2) culture conditions. Hypoxia rescue experiments were carried out by growing both NM and HYP cells in DMEM supplemented with 10% FBS under normal culture conditions (37ºC with 5%CO2 and 21%O2) for 3 or more days.

## Viral vector constructs and production

The ERT2 domain from the pCMV-CRE-ERT2 (Feil et al., 1996) was amplified by PCR and cloned into a modified FUGW lentiviral vector backbone (Lois et al., 2002). NTCF was then amplified by PCR from the pCDNA3.1-NTCF-NLS (van de Wetering et al., 2002a) and cloned in frame upstream of the ERT2 from the FUW-CMV-ERT2. The final FUW-CMV-NTCF-ERT2 vector is bicistronic and possesses an IRES sequence downstream of the ERT2 followed by a puromycin resistance cassette for selection of transduced cells. Viral production was accomplished in 293T producer cells (ATCC) by transient transfection of the lentiviral construct with packaging plasmids pCAG-RTR2 (Rev-expressing), pCAG-VSVG (Envelope) and pCAG-KGP1R (Packaging) (Henawa, Mol Ther, 2002) and linear polyethylenimine (PEI) (Polysciences Inc. 23966, Warrington, USA) as the transfection reagent. Viral supernatants were collected at 48h and 72h post-transfection, filtered using 0.45µm PVDF filters, supplemented with polybrene (8µg/mL) (Sigma H9268, St Louis, MO) and used to transduce target cells by O/N incubation. Infection efficiency of target cells was usually >90%, as determined following puromycin selection (2µg/mL, Invivogen, Toulousse, France). Cell lines stably expressing the corresponding constructs were maintained under antibiotic selection thereafter.

## Luciferase reporter assay

Beta-catenin/TCF activity in CRC cell lines was determined using TCF reporter plasmids pTOPflash and negative control pFOPflash as previously described (Morin et al., 1997). Briefly, pTOPflash and pFOPflash plasmids were transfected with empty vector (pCDN3.1) and transfection control Renilla luciferase (pRL-TK; Promega, Madison, WI) at a total concentration of 1µg/mL. Transfection was performed in the presence of linear PEI as our transfection reagent at a DNA/PEI mass ratio of 1:3. Cells were lysed in 1X Passive Lysis Buffer (Promega E194A) and luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega E641A and E195A). Firefly luciferase (TOP and FOP) activity was normalized to the transfection control pRL-TK to attain relative light unit values and the ratio of TOP/FOP was calculated. Activity was measured at 36h post-induction with 4-hydroxytamoxifen (4OHT, Sigma H7904) or treatment with vehicle (100% ethanol), unless otherwise stated. Luciferase assays were performed in triplicate and representative assays are shown.

## Microarray platforms

Biological replicates of all 11 CRCC expressing either the NTCF-ERT2 or ERT2 (44 microarrays) were used in the study, pooled expression profiles from three patients and the primary tumor sample SP5 were profiled on the HG-133 plus 2.0 microarray platform (Affymetrix). (CRC-SC) stem cell samples SP9 and SP42 and all hypoxia array expression profiles were produced on the human Prime View microarray platform (Affymetrix). All microarrays were processed in our core facility (IRB Transcriptomics) using standard techniques.

## Generation of WNT ON and WNT OFF gene expression signatures

SW948, SW403, LS174T, COLO205, SW480, SW620, SW48, CACO2 and DLD1 CRC cell lines stably expressing either the ERT2 or NTCF-ERT2 were seeded at 30% confluence and treated with 4OHT (Peprotech; 1uM $mL^{-1}$) for 36 hours. Gene expression profiles were measured in duplicate using HG-U133 plus 2.0 Affimetrix microarrays. RMA algorithm was used (Irizarry et al., 2003) to do background correction, median polish summarization and quantile normalization. The WNT ON signature (WNT ON) was obtained by selecting genes that possessed at least a 2.5-fold down-regulation when comparing the ERT2 to NTCF-ERT2, was consistent in 2 of 9 CRCC lines and possessed a pvalue < 0.05. The WNT OFF signature (WNT OFF) was obtained by selecting genes that possessed at least a 2.5-fold up-regulation when comparing the ERT2 to NTCF-ERT2, was consistent in 2 of 9 CRCC lines and possessed a

pvalue < 0.05. To build the WNT ON and OFF programs we excluded two cell lines: i. SW480 because they were derived from the same patient as SW620. In this way we avoided overrepresentation of WNT genes present in that particular tumor. ii. Lovo as this cell line displayed very low levels of beta-catenin/TCF activity.

**Assessment of WNT ON/OFF signatures in human CRCC lines and primary tissue**

The heatmap of WNT ON and OFF genes was obtained with Spearman distance and complete linkage for columns and Euclidean distance and complete linkage for rows. Genes that up or down-regulated at least 1.5-fold in 1of 9 CRCC lines are shown in the heatmap. The bars below the heatmap identify WNT ON and OFF signature genes.
To control for the specificity of the WNT blockade (NTCF-ERT2/ERT2) we analyzed significant gene expression changes in each of the 9 CRCC lines used above and two control cell lines LOVO and CO115, that possessed little of no WNT pathway activity, respectively. The comparison between NTCF-ERT2 and ERT2 (Supplementary Figure S1C) of the CRC cell line microarrays was performed by computing fold changes and pvalues (through a moderated t-test as implemented in the limma package from Bioconductor) if the fold changes were different from zero. P-values have been adjusted for multiple testing with the Benjamini and Hochberg method (Hochberg, 1995) to control the false discovery rate (FDR) at a level of 0.05.
For normal human colon (one EPHB2-high and one EPHB2-low for 3 different patients) and primary tumor samples SP5 (one EPHB2-high, one EPHB2-low and one EPHB2-neg), SP9 (one EPHB2-high and one EPHB2-low) and SP42 (one EPHB2-high and one EPHB2-low) tumor cell populations were corrected for probe intensity biases. Gam approximation was used to estimate and remove biases in mean and deviation.

**Generation of WNT ON HR<-1 and WNT ON HR>+1 gene expression signatures**

Patient information has been downloaded from GEO (Barrett et al., 2007). Available annotated clinical data for GSE17537 (55 patients) and GSE14333 (290 patients) included AJCC staging, age, gender and disease free survival intervals. In order to remove systematic biases, the expression measurements were converted to z-scores for all genes prior to merging these datasets. The latter merged datasets are referred to as the METAcohort from here onward. Stage 4 patients were removed from this cohort for all analysis except for the correlations provided in Table S2. After removing stage 4 patients the

METAcohort possessed 264 patients. A second patient cohort GSE33113 (89 patients) was also used. Expression measurements were converted to z-scores when the cohort was used to build a score with a gene signature. The latter cohort was analyzed using linear scale, as previously published (de Sousa et al., 2011). A third patient dataset was downloaded from TCGA (CGARN, 2008). Pooled human methylation data at level 3 (as defined in TCGA) from COAD (colon adenocarcinoma) and READ (rectum adenocarcinoma) was used to identify genome wide methylation patterns. There are a total of 222 patients with in this dataset. One probeset was selected for every gene in both, the METAcohort and the GSE33113 data. For every gene the selected probeset was the one with highest interquartile range across all samples as implemented in the nsFilter function form bioconductor's package genefilter.

Relapse hazard ratios for both METAcohort and GSE33113 cohort were performed with univariate Cox proportional hazards models (Cox, 1972). P-values were obtained with likelihood ratio-tests. WNT ON HR<-1 gene signature. These are the probesets in the WNT ON gene set that have a negative correlation with recurrence in the GSE33113 data. To define the list of negatively correlated probesets we required a probeset to have a negative hazard ratio for recurrence and a statistically significant pvalue < 0.05 for the test. In case of having more than one probeset of the same gene we kept the one with highest interquartile range in the GSE33113 data. A gene is considered to be regulated through methylation when 5% of the samples have methylation values above 0.25. WNT ON HR>+1 gene signature. These are the probesets in the WNT ON gene set that have a positive correlation with recurrence in the GSE33113 data. Methodology is the same as the previous signature but instead of requiring the hazard ratio to be negative we require it to be positive.

Pearson correlation estimates and P-values of the correlation between the average WNT ON HR<-1 and every gene (not in WNT ON HR<-1) were obtained for every cohort (Figure 2C). P-values were adjusted for multiple testing with the Benjamini and Yekutieli method (Benjamini and Yekutieli, 2001).

**Generation of Hypoxia gene expression signatures**

The CRC cell line LS174T and primary spheroid CRC stem cell (SP5) were cultured under hypoxic (1%O2) (HYP) or normal (21%O2)(NM) conditions for a period of 7 consecutive days. Gene expression profiles were measured by comparing 3 NM and 5HYP using Primeview Affymetrix microarrays. RMA algorithm was used (Irizarry et al., 2003) to do background correction, median polish summarization and quantile normalization. The hypoxia response

signature (HYP-RS) was obtained by selecting genes that possessed at least a 2-fold up-regulation in HYP compared NM. To define the list of up-regulated genes we computed moderated t-test statistics (Smyth, 2004), as implemented in the Bioconductor library limma. We then computed the probability that each gene is differentially expressed by fitting a semi-parametric partial t density (Rossell et al., 2008),and obtained a list of differentially expressed genes by controlling the Bayesian FDR below 5% (Müller et al., 2004). Genes that were considered to be WNT genes were removed from this signature. HYP-RS (primary-tumor organoids) was analysed as the latter hypoxia response signature.

**Subcutaneous and orthotopic tumor xenograft models**

CRC cell lines expressing ERT2 and NTCF-ERT2 were injected subcutaneously into the left and right flanks respectively of 5 to 6 week old Swiss nude mice (Jackson Labs) mice and followed for the periods described. This allowed direct comparison of control and WNT abrogated xenografts. For tumor initiation experiments, mice were fed withTAM400 diet (Teklad) 1 week prior to CRC cell injection and continued throughout the duration of the experiment. CRC cell lines were treated at time 0h with 4OHT prior to injection into mice. 4OHT was injected into mice (i.p.) every three days, thereafter. In tumor progression experiments, untreated CRC cell lines were injected into mice and allowed to engraft. Once tumor appearance was confirmed by palpation, 4OHT was injected into mice (i.p.) every three days and mice were immediately placed on TAM400 diet. Tumor dimensions were measured by caliper throughout the study to assess tumor volumes. All experiments were performed with At least 5 animals. Statistical significance was determined by assessing growth rate. The growth rate is the slope of the regression line (with tumor volume as response variable and days as explanatory variable) fitted with the least squares approach. The plotted growth rate is the log fold change between ERT2- average growth rate and NTCF-ERT2- average growth rate.
To identify hypoxic regions, Intra-caecum injections were performed with 1x105 cells in 6–8-week-old Balb/c mice. Mice were anaesthetized and injected retro-orbitally with 60 mg/kg pimonidazol hydrochloride (Hypoxyprobe-1, Chemicon) 30 minutes before sacrificing and tissues were fixed in formalin. Alternatively, when anaesthetizing equipment was not available Hyproxyprobe was injected i.p. 1hour before sacrificing. To identify hypoxic regions, intraocular administration of 60 mg/kg pimonidazol hydrochloride (Hypoxyprobe-1, Chemicon) was performed Orthotopic mouse models were approved by the animal care and use committee of the Barcelona Science Park (CEEA-PCB) and the Catalan Government (P18-R5-09).

## Tumor cell disaggregation and FACS sorting of EPHB2 populations

Primary tumor samples were obtained from Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). Samples SP5, SP9 and SP42 were implanted subcutaneously into NOD/Scid mice and grown as xenografts prior to their analysis. EPHB2 antibody was used to stain and purify high and negative tumor cell populations by FACS. A detailed protocol for tumor disaggregation and tumor cell purification was previously published by our group (Merlos-Suarez et al., 2011).

## Human colon crypt disaggregation and FACS sorting of EPHB2 populations

Primary tissue samples were obtained from Hospital del Mar (Barcelona). . EphB2 antibody was used to stain and purify high and low tumor cell populations by FACS. A detailed protocol for Human colon crypt disaggregation and CoSC cell purification was previously published by our group (Jung et al., 2011).

## Cell cycle assay

BrdU incorporation assays were performed as previously described (van de Wetering et al., 2002a). In Brief, $3x10^6$ cells expressing ERT2 or NTCF-ERT2 constructs were seeded in $56cm^2$ dishes (Corning 430167 NY, USA) in the presence of 1µM 4OHT or vehicle for a period of 36h. Cells were incubated with BrdU (Sigma B5002) between 20-60min, and then fixed with 70% ethanol. Nuclei were isolated and incubated with anti-BRDU-FITC (1:50) (Becton Dickinson, San Jose, CA). Cell cycle profiles were obtained by FACS. Ratios between the S-phase of ERT2 and NTCF-ERT2 are shown for corresponding cell lines. Representative assays are shown.

## RNA extraction, cDNA synthesis, microarray hybridization and real time-PCR

Total RNA was extracted from samples using TRIzol (Invitrogen 15596-018) followed by RNA column purification using the RNeasy kit (Qiagen 74106). Briefly, CRC cell lines were scraped from cell culture dishes (Costar) and homogenized by pipetting in TRIzol solution. After phase separation with chloroform, the upper aqueous phase was then mixed with RLT lysis buffer according to the standard RNeasy protocol provided by the manufacturer, with the exception of two additional washes with the RPE buffer to remove trace amounts of phenol.

RNA was quantified using a Nanodrop spectrophotometer and assessed for quality using a Bioanalyzer (Agilent) and used as template to hybridize to HG-133 plus 2.0 microarrays (Affymetrix) in our core facility (IRB Transcriptomics) using standard techniques. Only intact RNA with RIN scores between 9.5 and 10 were hybridized.
For qRT-PCR, the High Capacity cDNA Archive Kit (Applied Biosystems) was used to reverse transcribe 1 ug of purified RNA to cDNA according to the manufacturer's instructions. All qRT-PCR reactions were performed using Taqman probes (Applied Biosystems) and Taqman Universal PCR Master Mix (Applied Biosystems). The Taqman probes used in this study are the following: APCDD1 (Hs00537787_m1), ASCL2 (Hs00270888_s1), AXIN2 (Hs00610344_m1), EPHB2 (Hs00362096_m1), PPIA (Hs99999904_m1), CD44 (Hs00153304_m1), CXCR4 (Hs_00237052_m1), PTPRO (Hs00243097_m1). StepOnePlus Real-Time PCR and ABI Prism 7900 Sequence Detector Systems were used to carry out the qRT-PCR reactions in clear optical 96-well reaction plates with optical covers, according to manufacturer's instructions. Gene expression levels were normalized using the endogenous control PPIA for each sample and differences in target gene expression were determined using SDS 2.4 or StepOne 2.2 plus software. Error bars represent standard deviation of samples performed in triplicate.

**Immunohistochemistry (IHC)**

Xenograft tissue from intracaecum injections was fixed in 10% neutral buffered formalin (Sigma) o/n and embedded in paraffin blocks. Immunostainings were performed using 4 um tissue sections according to standard procedures. Briefly, antigen retrieval was performed, samples were blocked for 1h at RT, and then the primary antibody was incubated in the blocking solution. Slides were washed with PBS and a corresponding secondary antibody was incubated with the sample for 45 min at RT. Samples were developed using DAB, counterstained with haematoxylin and mounted. Primary antibody details, dilutions, retrieval, blocking and incubation conditions are outlined in the table below. . Double immunohistochemistry was performed by revealing secondary antibodies (Brightvision) with ImmPACT VIP and DAB peroxidase Substrates (Vector Labs).

| **Primary antibody** | **Manufacturer and Reference** | **Dilution** | **Retrieval, blocking, primary antibody incubation** |
|---|---|---|---|
| Mouse anti-KRT20 | Dako (M7019) | 1:200 | Tris-EDTA pH9, boil 20 min, BSA 1%, O/N at 4ºC |
| Rabbit anti-FABP1 | Sigma (HPA028275) | 1:500 | Citrate pH6, autoclave 20 min, BSA 1%, O/N at 4ºC |

| | | | |
|---|---|---|---|
| Mouse anti-Beta-catenin | SIGMA (7207) | 1:100 | Citrate pH6, autoclave 20 min, BSA 0.05%, 20 min at RT |
| Goat anti-Epcam | R&D (AF960) | 1:50 | Citrate pH6, autoclave 20 min, BSA 1%, O/N at RT |
| Goat anti-EphB2 | R&D (AF467) | 1:200 | Citrate pH6, autoclave 20 min, BSA 0.05%, O/N at 4ºC |
| Rabbit anti-MYC | Upstate (06-340) | 1:500 | Tris-EDTA pH9, boil 40 min, BSA 1%, 72 h at 4ºC |
| Mouse anti-BRDU | Beckton (347580) | 1:100 | Citrate pH6, boil 20 min, BSA 0.05%, 1h at RT |
| Mouse anti-LAMC2 | Millipore (MAB19562X) | 1:100 | Tris-EDTA pH9, boil 20 min, BSA 1%, o/n at 4ºC |
| Rabbit anti-CXCR4 | Epitomics (3108-1) | 1:100 | Tris-EDTA pH9, boil 40 min, DAKO diluent, o/n at 4ºC |
| Rabbit anti-GLUT1 | Millipore (07-1001) | 1:5000 | Citrate pH6, boil 20 min, BSA 5%, O/N at 4ºC |

**Immunofluorescence (IF)**

The following primary antibodies were used; Mouse anti-KRT20 (Dako, 1:100), Rabbit anti-beta-catenin (Santa Cruz,SC-7199, 1:100), Goat anti-EPCAM (Sigma, 1:50), Mouse anti-LAM5 (Milipore, 1:100) and mouse anti-EPHB2 (R&D, 1:100). Secondary anti-mouse 649 DyLight conjugated, anti-goat xRed conjugated and anti-rabbit FITC conjugated (Jackson Immunoresearch, 715-495-150; 705-295-147; and 711-485-1521, respectively) were all of Donkey origin and were used at a 1:400 dilution for 1h at RT.

Briefly, LS174T cells were cultured on treated glass slides (Costar) at 5% confluence (5.000 cells/cm$^2$) and allowed to grow for 7 days in normal (NM) oxygen conditions (21% O2) or under hypoxic (HYP) conditions (1%O2). Cells were fixed in 4% PFA for 10 min at RT, incubated in PBS containing 20mM glycine for 10 min, permeabilized in PBS containing 0.25% Triton-X 100 for 20 min, blocked with PBS containing 1% BSA (Sigma) for 30 min. Immunostaining on paraffin embedded tumor sections were performed using 4 um tissue sections according to standard procedures. Samples were then blocked with DAKO diluent (DAKO, K8006) for 30 min and incubated with the primary antibody overnight at 4°C followed by incubated with the secondary antibody for 1h at RT hour in the dark. Nuclei were stained with DAPI (Sigma, D9542) at 0.1µg/ml for 10 min. Slides were mounted in Vectashield (Vector labs H-1000 Burlingame, CA) or DAKO (DAKO, S3023) mounting medium.

## Tumor initiation assays

EPHB2 tumor populations were defined as follows: EPHB2$^{high}$ cells were considered the brightest 10% of the EPHB2 population. EphB2$^{low}$ cells were the 20% whose mean fluorescence intensity (MFI) was 4 times lower than the EphB2high population. The EPHB2$^{neg}$ population was considered as the 20% of cells which had a 30 to 40 fold lower MFI than the EPHB2$^{high}$ fraction. These three tumor cell populations were sorted using a FACS Aria 2.0 cell sorter (BD). A detailed protocol for tumor initiation was previously published by our group (Merlos-Suarez et al., 2011).

## Immunoblotting

Briefly, protein extracts were obtained by lysing cells in 1mM EDTA, 1mM EGTA, 1% SDS supplemented with fresh Protease inhibitor solution (Sigma P8340), homogenized by pipetting and boiled for 10 min. Protein concentration was measured using the Bradford Protein Assay kit (Bio-Rad). Cell lysates containing 20µg of protein were separated by SDS gel electrophoresis and transferred to PVDF membrane (Millipore). Primary antibodies were incubated O/N at 4ºC with 5% milk in TBS-Tween 0.1% (blocking solution). The primary antibodies used were Goat anti-EPHB2 (R&D, AF467, 1:1000) and Rabbit-anti-ACTIN (Sigma, A5060, 1:10000). Horseradish peroxidise (HRP) conjugated secondary antibodies (Pierce) were routinely used at a 1/5000 dilution and incubated for 1 hour at RT in blocking solution. Membranes were washed in TBS-Tween 0.1%. Immunocomplexes were detected using ECL kit (GE healthcare) according to the manufacturer's instructions.

## Quantification of budding tumor cells at invasion fronts

Over 100 photographs of invading cells from 10 tumor samples possessing strong nucelear beta-catenin staining were assessed for the colocalization of nuclear beta-catenin (BCAT) and cytokeratin 20 (KRT20). Double immunofluorescence was performed on human primary colorectal tumor samples against KRT20, BCAT and counterstained with DAPI. Only Budding cells possessing nuclear BCAT were included in the analysis and placed into 3 categories as follows: 1) KRT20 high, 2) KRT20 low and 3) KRT20 negative. The percentages of cells that fell into each category were calculated by averaging the number of cells that fell in each of the 3 categories.

# CHAPTER 5

# GENERAL DISCUSSION AND FUTURE PROSPECTS

## The GATA6 transcription factor converges on BMP signaling to regulate AdSC self-renewal

It is well documented that CRC occurs as a result of the stepwise acquisition of genomic aberrations (Fearon and Vogelstein, 1990). The most prevalent of these mutations is aberrant activation of the WNT signaling pathway, which is present in > 90% of sporadic colorectal cancers (Cancer Genome Atlas, 2012). Recent work has identified LGR5 as an ISC restricted WNT target gene. LGR5(+) cells are located at the base on intestinal crypts, self-renew and give rise to all cell lineages of the intestinal tract throughout life (Barker et al., 2008; Barker et al., 2007). These cells have been pinpointed as the cell of origin of intestinal cancer (Barker et al., 2009). Remarkably, Adenomas possess LGR5(+) cells localized towards the base of tumor glands that behave as adenoma stem cells (AdSCs) whereas the rest of tumor cell types represent differentiated progeny of these AdSCs (Schepers et al., 2012).

Great efforts have been focused on understanding the molecular pathways and regulators of tumor stem cells in colorectal pre-malignant and malignant lesions. The hierarchical organization of these lesions has drawn significant attention to the potential of specifically targeting CRC-SC. Thus, understanding how hierarchies are imposed and how they regulate pre-malignant and malignant stem cell numbers is of significant relevance in the field. In the present study we have used a bioinformatic approach to identify regulators of the tumor stem cell phenotype. We were able to construct a gene regulatory network centered on ISC and AdSC related genes. A shRNA screen of the nodes in the ARACNe network led to the identification of the transcription factor GATA6 as a key player in the regulation of cellular hierarchies in colonic adenomas. We show that *GATA6* and *LGR5* expression are intimately related in CRC cells and in murine pre-malignant lesions. We further demonstrated by ChIP-seq that this regulation likely occurs through direct binding of GATA6 to the *LGR5* proximal promoter. In agreement with our findings, previous work has also shown that modulation of *GATA6* expression directly affects *LGR5* expression in differentiating embryonal carcinoma P19 cells (Alexandrovich et al., 2008). Moreover, these same authors also demonstrated that GATA6 bound to the *LGR5* proximal promoter in African green monkey kidney fibroblast-like COS-7

cells. This suggests that GATA6 has a conserved role in regulating *LGR5* expression in different systems.

In addition to controlling *LGR5* expression, we also show that GATA6 regulates the expression of another ISC gene *Cldn2*. We demonstrate for the first time that the protein levels of CLDN2 are highly restricted to ISC and AdSC in the base of colon crypts. Microarray analysis has shown that CLDN2 was enriched in mouse (Munoz et al., 2012) and human (Jung et al., 2011) ISCs. Moreover, CLDN2 expression closely reflects that of LGR5 in adenomas. CLDN2 is a tight junction protein restricted to the cell surface, making it a novel and attractive ISC marker for work focused on isolating and characterizing normal and tumor stem cells. Indeed, future work could explore the use of CLDN2 antibodies to enrich for human CoSC and tumor initiating cells (TICs) in later stages of CRC. Few commercial antibodies are available to enrich these populations and the published stem cell markers CD133 (O'Brien et al., 2007; Ricci-Vitiani et al., 2007), CD44 (Dalerba et al., 2007b), CD166 (Dalerba et al., 2007b) and EphB2 (Merlos-Suarez et al., 2011) display broader expression patterns than CLDN2. The identification of this GATA6 regulated stem cell markers opens up new avenues for future studies of stem cells from the intestinal tract.

In addition to being expressed in ISCs and AdSC, CLDN2 has recently been shown to be upregulated in an aggressive breast cancer cell line that preferentially metastasises to liver. Also, it was shown that CLDN2 is expressed in 100% of breast cancer liver metastases (Tabaries et al., 2011). The tropism of metastatic breast cancer cells to the liver was related to CLDN2 function by allowing breast cancer cells to adhere to hepatocytes (Tabaries et al., 2012). Interestingly, a recent study showed that *GATA6* knockdown negatively affected CRC migration and invasion, whereas aberrant GATA6 expression in CRC was associated with metastasis and poor overall survival (Shen et al., 2013). These observations may have important implications to understand metastatic CRC and warrants further investigation into the link between CRC-SC and liver metastasis.

GATA factors have been shown to have a number of redundant functions as they bind equivalent DNA sequences. Also, several GATA isoforms are often co-expressed during development and in numerous adult organs (Merika and Orkin, 1993; Peterkin et al., 2007). We confirmed this redundancy by demonstrating that overexpression of either *GATA4* or *GATA6* can affect the regulation of ISC specific genes *LGR5* and *CLDN2*. GATA5 is also capable of modulating the expression of these two genes (data not shown). Thus, it is not surprising that the phenotypic consequence of *Gata6* loss is most pronounced in the colon, as it is the only GATA factor expressed in this part of the

gastrointestinal tract. As such, it would be interesting to conditionally delete all GATA factors from the entire intestinal epithelium in an *Apc* null background; yet, such experiments would require a lengthy and complex mouse breeding schemes.

GATA factors also have divergent roles depending on the cell type and spacio-temporal expression, such as that seen during embryonic development. In contrast to intestinal adenomas, *Gata6* loss in the airway epithelium results in the amplification of bronchioalveolar stem cells (Zhang et al., 2008). Furthermore, while we have shown that loss of *Gata6* in colon adenomas results in decreased AdSC numbers and increased BMP ligands (including *Bmp2*), its loss in embryonic stem (ES) cells results in decreased *Bmp2* expression and apoptosis (Rong et al., 2012). Interestingly, GATA6 was shown to bind the proximal promoter in ES cells in the latter study, whereas we failed to find GATA6 bound to this region in CRC (data not shown). This observation suggests that GATA6 specificity is at least in part specified by the epigenetic landscape of the cell type in question. Furthermore, it has recently been suggested that GATA factors perform DNA looping at promoters and enhancers. This feature allows GATA factors to bridge two DNA fragments using their dual zinc finger activity, implying an ability to exert long range DNA regulation (Chen et al., 2012). Taken together it is not surprising that GATA factors have such divergent gene regulatory roles when considering the diverse epigenetic landscape found in different cellular contexts.

BMP signaling has been documented to counteract the WNT gradient in intestinal crypts (Kosinski et al., 2007). In fact, BMP inhibitors are required for the long-term culture of normal ISC from mice and humans (Jung et al., 2011; Sato et al., 2009). Moreover, in the absence of BMP antagonists (e.g. NOG or DMH1), a decrease in *Lgr5* expression has been observed in tumor organoid cultures (Sato et al., 2011), highlighting the essential role of BMP signaling in the control the stemness. However, clonogenicity in the later study was shown to be unperturbed over serial passaging, implying that a particular BMP threshold must be reached to negatively impact on tumor organoids. This fact is evidenced by irreversible differentiation of tumor organoids when culture medium is supplemented of with rBMP4 (Farrall et al., 2012). In agreement with these findings we show that supplementation of culture medium with rBMP2 or rBMP4 or deletion of *Gata6* in tumor organoids results in differentiation and consequent loss of AdSC self-renewal as evidenced by loss of tumor organoid clonogenicity. Importantly, we demonstrate that the overt phenotype resulting from deletion of *Gata6 in vitro* can be rescued by supplementing culture media with the BMP antagonists DMH1 or NOG, establishing the transcription factor GATA6 as a regulator of BMP signaling.

*In vivo* it has been shown that BMP signals are largely absent from intestinal crypts while the BMP pathway agonist BMP4 has been reported to be confined to the mesenchyme in the normal intestinal tissue (Haramis et al., 2004; He et al., 2004). More recently *Bmp4* has been shown to be expressed in the epithelial compartment of adenomas from the small intestine (Farrall et al., 2012). As a result, BMP4 was proposed to act in an autocrine loop in adenomas to limit self-renewal and play a role in creating cellular hierarchies (Farrall et al., 2012). Indeed, our study confirms that the BMP pathway is significantly activated in adenomas, although in contrast to adenomas in the small intestine described by Farrall et al. (2012), we report this activation in the colon and show that BMP pathway activation is predominantly absent in the lower two-thirds of crypts and present in the upper third and in the bulk of the adenoma glands. This observation suggests the existence of niche located towards the base of adenoma glands that protect AdSCs from BMP signaling. Fittingly, it has been previously evidenced that the mesenchyme surrounding the crypt base contributes BMP antagonists (e.g. *NOG, GREM1/2*, and *CHRDL1*). These adenoma compartments are perturbed in *Gata6* null adenoma cells, as evidenced by numerous P-SMAD1/5/8 and ID1 positive cells located in the lower two-thirds of the adenoma crypts and glands. BMP pathway activation at the crypt base enforced by Gata6 deletion coincided with a significant decrease in tumor burden as well as a significant decrease in the expression of ISC/AdSC markers LGR5 and CLDN2. The phenotype of *Gata6* loss was more prominent in cultured tumor organoids since under these conditions no underlying mesenchyme (Kosinski et al., 2007) is present to counteract the autocrine BMP signals originating from tumor organoid cultures.

Another set of studies has shown that abrogation of BMP signals by conditional deletion of *Smad4* in an *Apc* $^{\Delta 1638}$ background increased tumor burden and WNT target gene expression (Freeman et al., 2012). Interestingly, these authors also show that BMP signaling controls beta-catenin transcription in a *Smad4* dependent fashion *in vivo* and *in vitro* using colorectal cancer cells. It is interesting to note that BMP4 was found to be a WNT target gene in colorectal cancer cell lines (Kim et al., 2002; Scholer-Dahirel et al., 2011; van de Wetering et al., 2002b), which suggests a possible regulatory interaction between the two pathways. This link is evidenced in our study by the fact that treatment with recombinant BMPs or *Gata6* deletion in tumor organoids resulted in massive upregulation of WNT inhibitors *Apcdd1*, *Dkk2*, *Dkk3*, *Nkd1*, *Notum*, *Wif1* and *Znrf3*. Likewise, *GATA6* shRNA knockdown in CRC cells resulted in the upregulation of *DKK1*, *NKD1* and *APCDD1* (data not shown). Fittingly, one of the downstream BMP effectors, SMAD1, has been shown to control the expression of WNT inhibitor WIF1 in lung development (Xu et al., 2011).

Considering that our model systems have mutational activation of WNT signaling as a result of downstream mutations in the pathway (APC or beta-catenin mutations) it seems unlikely that these WNT inhibitors play a major role in suppressing tumorigenesis, as all of them have been reported to act upstream of APC. Nonetheless, it has been demonstrated that AXIN2 and APCDD1 have a negative impact on beta-catenin/TCF activity in WNT mutant CRC cell lines (de Sousa et al., 2011). Thus, in future work it may be worth exploring the impact these inhibitors have alone or in combination on AdSC self-renewal to assess their role on canonical and non-canonical WNT pathway activities. These observations suggest that GATA6 is functioning not only to limit BMP signals, but perhaps also by limiting WNT signals.

Further evidence linking BMP and WNT pathways to GATA6 expression come from ChIP-seq experiments performed on CRC cells in our study. We show that GATA6 does not bind to the TSS of BMP4, but instead binds to distant putative BMP4 enhancer regions. One such GATA6-bound enhancer region maps to SNPs in linkage disequilibrium with a SNP (rs1957636), which has been previously associated to CRC susceptibility in GWAS studies (Tomlinson et al., 2011a). Interestingly, analysis of ChIP-seq data suggested that the beta-catenin/TCF4 complex is also bound to this region. Our data indicate that beta-catenin/TCF4 and GATA6 inversely regulate *BMP4* expression. Accordingly, both factors bind differentially to *BMP4* enhancer regions. We hypothesize that alterations, such as particular SNPs, that modulate the activity of these enhancers may impinge on the levels of BMP4 and therefore influence the expansion of AdSCs at the onset of tumorigenesis. Increased susceptibility to CRC associated to SNPs in this region may be the result of changes in the transcriptional circuit regulated by GATA6.

Therefore, future work should broadly explore the interplay between beta-catenin/TCF and GATA6 to dissect what factors may play a role in their expression and activities. An attractive example would be analysis of enhancer derived RNAs (eRNAs), as they have been recently shown to specifically and functionally act at enhancer regions where they are transcribed to aid in mediating looping of chromatin to the TSS sites of adjacent genes (Lam et al., 2013; Li et al., 2013). These *cis* acting eRNAs could further elucidate the complex regulation GATA6 has in tissues such as the colon where the vast majority of ChIP-seq peaks were found to be located at enhancer regions. Tissue specific enhancer activities have been previously reported to exist implying selectivity of enhancers depending on the cellular context (Visel et al., 2009). For example, unlike *GATA6* KD CRC cell lines LS174T, *Gata6* null embryoid bodies have decreased levels of BMP2 and BMP4 expression (Rong et al., 2012). This is in line with the divergent role GATA6 has in promoting

(Belaguli et al., 2010; Lin et al., 2012) or inhibiting (Cai et al., 2009; Lindholm et al., 2009) tumorigenesis in different tissues.

Recent work has identified the genome reorganizer SATB1 as a positive regulator of GATA3 in T-helper type 2 cells and a direct competitor of TCF4 for beta-catenin recruitment (Notani et al., 2010). Interestingly, both known SATBs (SATB1 and SATB2) are directly connected to both GATA6 and LGR5 in the predicted ARACNe ISC gene network. SATB1 has been described to promote cancer formation in different tissues and its expression often correlates with the disease progression (Mir et al., 2012). Furthermore *Satb1* depletion in hematopoietic stem cells (HSC) has been shown to affect self-renewal and accelerate lineage commitment (Will et al., 2013) somewhat paralleling the effects seen upon *Gata6* deletion in colon AdSCs. Further elucidation the roles of these factors in relation to GATA6 may shed light on complex gene regulation switches between tumor stem cells and their progeny.

BMP ligands have been shown to impinge growth of CRC cell lines, which is in line with the idea that the BMP pathway inactivation may play a role in later stages of CRC (Beck et al., 2006; Hardwick et al., 2008; Hardwick et al., 2004). In fact, a recent study showed that BMP4 not only induced differentiation of colorectal cancer stem cells (CRC-SC), but also increased their sensitivity to chemotherapy in mouse xenografts (Lombardo et al., 2011). In line with these observations, loss of BMP activation as determined by P-SMAD1/5/8 staining has been associated with the adenoma to carcinoma transition (Kodach et al., 2008b). This suggests that continued GATA6 expression may be a determining factor to maintain BMP signals low and allow disease progression. Thus, it is tempting to speculate that epithelial derived BMP signaling in adenomas serve as a molecular brake to impede both growth and progression of the disease by limiting the AdSC/CRC-SC numbers. This is in line with the observation that adenomas generally do not possess defects in BMP signaling, whereas 70% of CRCs have acquired BMP pathway mutations (Kodach et al., 2008a) implying that loss of BMP signaling marks the adenoma to carcinoma transition. In fact this transition is often accompanied with TGF-beta loss when the shared co-SMAD, SMAD4, is inactivated.

To a large extent, the role of the BMP pathway in CRC progression has largely been ignored in the literature as it continues to remain in the shadow of the more well described canonical TGF-beta pathway. Nonetheless, it is clear that BMP signaling between epithelial and stromal compartments in the intestine are critical for disease, as the simultaneous disruption of BMP signaling in both compartments, resulted in polyposis and even carcinogenesis (He et al., 2004; Kim et al., 2006). Likewise compound *Smad4*/*Apc* mutants also developed

more aggressive tumors than those mice that carried mutational inactivation of *Apc* alone (Takaku et al., 1998). The mounting evidence that BMP signaling plays a role in tumor hierarchies and CSC/AdSC clonogenicity warrants further investigation into this extensive branch of the TGF-beta superfamily. Moreover the newly found link between GATA6/BMP and WNT signaling pathways shown in this work may have important implications in other cancer types and may aid in better elucidating the functional diversity of GATA family members.

## A Snail independent novel type of EMT is specified by GATA factors and is conserved from flies to mammals

In cancer much like in development, classical EMT is triggered by transcriptional inhibition of adherens junction proteins, most notably E-cadherin (*CDH1*) (Thiery et al., 2009). This is typically accomplished by direct or indirect binding of transcriptional repressors to the E-box consensus sequences in the *CDH1* promoter (Batlle et al., 2000; Cano et al., 2000). Ultimately this transition leads to the conversion of a relatively static to highly motile cellular state. The presented work adds new players, Serpent/GATA4/6 (GATA factors), to a growing list of EMT inducers. Importantly, these GATA factors function independently of Snail genes (Barrallo-Gimeno and Nieto, 2005) and do not result in alternations in *CDH1* RNA and protein abundance, but instead in its cellular localization (i.e. translocation from membrane to cytoplasm). Among the EMT inducing transcription factors that indirectly effect *CDH1* levels, FOXC2 stands out as one of the most similar to GATA factors as it has been shown to have no effect of *CDH1* transcription and has been described to relocalize CDH1 protein from the membrane to the cytosol. However, unlike the GATA factors described herein FOXC2 imposes a strong decrease of the CDH1 protein level (Mani et al., 2007). Moreover, FOXC2 is typically expressed in cells which co-express direct *CDH1* repressors such as SNAI1, TWIST and Goosecoid (Yang and Weinberg, 2008).

Fittingly with our findings in the mammalian MDCK cells, GATA4 has been implicated in the regulation of a Snail independent EMT in endothelial-derived cells in the murine developing heart (Rivera-Feliciano et al., 2006). In this study, *Gata4* mutant endothelial cells could not form atrioventricular (AV) valves as they failed to undergo EMT. It would be interesting to see if the conditional deletion of *Gata6* has a comparable phenotypic outcome as that of *Gata4* in this context. Likewise a broader analysis of the role of GATA factors in EMT during development is warranted to decipher whether they allow plasticity of particular cell types during organ formation.

The evolutionarily conserved role GATA factors in the regulation of EMT through transcriptional downregulation of Crumbs is a major strength of this work. It would be interesting to explore whether this alternative route to EMT is highjacked in pathologies such as CRC, which may have important clinical implications. It is well established that EMT is implicated in invasion and metastasis and is embedded in the hallmarks of cancer (Hanahan and Weinberg, 2000, 2011). In CRC, budding cells at the invasion fronts of primary CRCs have been described to harbour undifferentiated stem-like cells (Brabletz et al., 2005). These invasion fronts with budding cells have been shown to correlate with poor prognosis (Prall et al., 2005; Tanaka et al., 2003) and local and distal metastasis (Kazama et al., 2006; Nakamura et al., 2005). It has been suggested that tumor buds may represent cells undergoing EMT. Yet, close inspection of these cells reveals that they do not downregulate E-cadherin nor do they gain expression of mesenchymal genes (data not shown). It is thus possible that migratory properties of CRC cells are acquired through transient modulation of adherent junction proteins. It is tempting to speculate that GATA6 may mediate these effects at invasion fronts. Two prominent markers of budding cells in CRC are Laminin subunit gamma-2 (LAMC2) and urokinase plasminogen activator (uPAR), which are co-expressed in these migratory cells (Pyke et al., 1994; Pyke et al., 1995). Of note, GATA6 has been shown to regulate *uPAR* gene expression and promote invasion in metastatic CRC cell lines through direct binding to Sp1 (Belaguli et al., 2010). The EMT inducing SDF1/CXCR4 axis has also been shown to regulate *uPAR* expression by increasing Sp1 binding to DNA (Huang et al., 2012). Interestingly, a very recent publication has reported that GATA6 expression correlated with poor overall survival in CRC and that its aberrant expression was associated to liver metastasis (Shen et al., 2013). Such overlapping regulatory roles certainly warrant further investigation into the possible cooperative roles between GATA6 and the SDF1/CXCR4 axis in CRC.

Seminal work has demonstrated that EMT endows somatic cells with cancer stem cell (CSC) properties (Mani et al., 2008; Morel et al., 2008), thus linking EMT to the cancer stem cell hypothesis (Dalerba et al., 2007a; Reya et al., 2001) and substantiating the original hypothesis that these budding cells may in fact be “migrating cancer stem cells” (Brabletz et al., 2001; Brabletz et al., 2005). Fittingly, our results highlight the dual role of GATA6 in both the EMT and maintenance of tumor organoid/AdSCs through BMP and WNT inhibitor repression. Strikingly, WNT inhibitor factor 1 (WIF1), which was massively upregulated upon deletion of *Gata6* in tumor organoids, has been shown to hamper migration, invasion and reverse EMT in prostate cancer cells (Yee et al., 2010). Conversely, using proteomics, mathematical models and *in vitro* invasion assays, a recent study has identified the BMP2/4/7 antagonist GREM1

as a secreted factor that may play a role in triggering the motility of tumor cells at the invasion front of CRC tumors (Karagiannis et al., 2013). Interestingly, they show that colorectal cancer cells lines responded to GREM1 in a chemotactic manner and that BMP7 was able to dramatically reduce this tendency, suggesting that this phenomenon is BMP dependent. Additionally, in the kidney, BMP7 has been described to induce MET (Zeisberg et al., 2005). Conversely, in the airway epithelium BMP antagonists GREM1 and NOG block cell migration while BMP2, BMP4 or BMP7 induced EMT (McCormack et al., 2013), again highlighting the tissue and cell specific role of GATA6.

It is well accepted that the acquisition of EMT in differentiated tumors is a key step leading to tumor cell dissemination (Thiery et al., 2009). However, primary tumors and their corresponding metastases in CRC possess comparable histological grades and often appear as well-differentiated tumors with similar proportions of differentiated and undifferentiated cells (Merlos-Suarez et al., 2011). This led to the proposal of a transient EMT that would be followed by reversion of the process (MET) at the site of metastasis (Brabletz et al., 2001). The transient EMT to MET we observed as a result of the GATA factor Srp in the PMG is reminiscent of this proposed model. Strikingly, recent work in a mouse model of skin cancer support this EMT-MET hypothesis by showing that a transient EMT induced at the primary tumor, as opposed to sustained EMT, leads to increased numbers of metastases despite comparable amounts of circulating tumor cells (Tsai et al., 2012). Future studies should indeed decipher the role GATA6 may have in the invasion fronts of CRC tumors. Furthermore, a central question that arises is the paradoxical role of GATA6 as a promoter of the self-renewal of AdSC at the onset of tumorigenesis and as a potential promoter of cell migration in advanced CRC stages. Epigenetics and the divergence in GATA6 transcription factor partners may explain this dichotomy. Finally, it is worth pointing out that as transcription factors are not easily amenable to drug targeting it will be worth pursuing whether up or downstream effectors as well as critical partners of GATA6 may be therapeutically targeted.

## A Subset of WNT/CRC-SC target genes are silenced in hypoxia and are associated with disease relapse

Aberrant activation of the WNT signaling pathway occurs in the vast majority of CRC tumors (Cancer Genome Atlas, 2012). Although a large number of WNT target genes have been identified by analyzing one or two CRC cell lines (van de Wetering et al., 2002a; Van der Flier et al., 2007) a comprehensive study was lacking. As such, we took the initiative of performing an in depth analysis of the beta-catenin/TCF gene program in 10 CRC cell lines. To this end, we

developed an inducible genetic system that mimics an ideal anti-WNT drug by blocking the interaction between nuclear beta-catenin and TCF/LEF. Importantly, we validated our WNT target lists (WNT ON and WNT OFF) by showing that the identified WNT ON target genes are enriched not only in CoSC but also CRC-SC. Of note, the WNT ON signature is not homogeneously expressed in CRC cell lines. This is the consequence of both the stringent cutoff used in the analysis and the large diversity of genes controlled by WNT signaling in CRC. While many key WNT targets previously identified were also found in our studies (e.g. *LGR5, ASLC2* and S*MOC2*), our thorough approach allowed for the identification of many novel and robust targets including the hedgehog co-receptor PTCH1, the drug transporter ABCB1 or the immunomodulator CADM1, all of which play important roles during development and cancer in other systems.

The impairment of nuclear beta-catenin has been shown (in our and other studies) to be fundamental in impairing tumorigenic potential of CRC cells *in vivo* (Scholer-Dahirel et al., 2011). However, in contrast to the latter study we show that the degree of sensitivity to WNT blockade can vary considerably between CRC cell lines. Notably, our study utilizing 10 CRC cell lines shows that expression levels of the WNT ON program are directly related to the dependency of pathway activity for grow *in vivo* (i.e. cells with highest levels of WNT ON gene expression are the most sensitive to WNT blockade, whereas those cells with the lowest expression are refractory to this blockade). This discovery has potentially far reaching clinical applications and suggests that molecular stratification of patients into two groups (i.e. those patients with the highest WNT ON gene expression and those with low levels) could identify those patients that would best respond to anti-WNT therapeutics, as an ever growing pipeline of such drugs are currently in clinical trials (Curtin, 2013; Takahashi-Yanaga and Kahn, 2010). This last point stresses the need to focus on those patients who would best benefit from therapeutic intervention in the era of personalized medicine.

The importance of aberrant WNT activation is well established as an initiating event in CRC tumorigenesis. Nevertheless, its role during disease progression has been less explored. It is well accepted that CRC-SC express elevated levels of WNT target genes and that these WNT high signatures endow CRC cells and CRC-SCs with self-renewal and tumorigenic potential (Merlos-Suarez et al., 2011; Scholer-Dahirel et al., 2011; van de Wetering et al., 2002a). Yet, upon analysis of the WNT expression program in patient cohorts we have confirmed the paradoxical observation made originally by de Sousa-Melo et al., 2011 that the silencing of a subset of the WNT target gene program predicts poor prognosis. Importantly, we show that an equivalent number of WNT ON

targets positively and negatively predict relapse, which provides a rational explanation for why the WNT ON signature does not hold predictive power for disease relapse. To explain this phenomenon de Sousa-Melo et al., 2011 identified a handful of apparently methylated WNT target genes that stratify those patients with recurrent disease, leading the authors to suggest that methylation was playing a role in fine tuning the WNT program during disease progression. Although we did observe methylation in some WNT ON genes that negatively associated with recurrence (WNT ON$^{HR<-1}$), these genes did not predict relapse any better than there non-methylated counterparts. Moreover, we observed comparative degrees of methylation in WNT ON genes that positively (WNT ON$^{HR>+1}$) and negatively (WNT ON$^{HR<-1}$) associated with disease relapse. Our analyses revealed that methylation of WNT ON genes in general strongly correlated with genome wide methylation rates despite the frequent methylation of a fraction of WNT targets. These observations suggest that, in fact, WNT genes are highly methylated in CRCs that display high rates of methylation (i.e. CpG island methylator phenotype (CIMP-high)) and that this phenomenon occurs independently of disease relapse.

We also report that expression of EPHB2, a marker for normal CoSCs (Jung et al., 2011) and CRC-SC (Merlos-Suarez et al., 2011) strongly correlates with WNT ON$^{HR<-1}$ in both patient tumor cohort. Our group had shown that *EPHB2* is silenced as the disease progresses and tumors become more aggressive to allow loss of cell compartmentalization and invasion (Batlle et al., 2005; Cortina et al., 2007). In fact, loss of EPHB2 strongly correlated with degree of malignancy (Batlle et al., 2005). The proposed mechanisms involved in EPHB2 silencing are promoter methylation and frame shift mutations (Alazzouzi et al., 2005), although we did not detect any methylation in the *EPHB2* promoter. Of note, others have claimed that *EPHB2* is not methylated but instead is regulated by the transcription factor REL (Fu et al., 2009).This data reinforces the idea that some WNT target genes are specifically silenced to allow disease progression despite sustained WNT pathway activation in late stage CRCs.

Another controversial study suggested that the concomitant acquisition of a Hedghog (HH) and an ES stemness program occurs in late stage metastatic CRC. This phenomenon is linked to silencing of the WNT ON program (i.e. TNM stage 1 and 2 to TNM stage 3 or 4). However, a very low number of patients were used to draw this conclusion (Varnat et al., 2010). Our results did not corroborate these findings. Rather, we found that most ES cell specific signatures (Ben-Porath et al., 2008) correlated very well with WNT ON genes that negatively predict relapse (WNT ON$^{HR<-1}$) in two independent CRC patient cohorts.

The unexpected identification of a large number of hypoxia related genes inversely expressed to the WNT ON$^{HR<-1}$ gene subset led us to explore the implication of hypoxia in this process. The culture of LS174T colorectal cancer cells under hypoxic conditions did not affect beta-catenin/TCF reporter activity, yet resulted in the downregulation of most of the WNT ON$^{HR<-1}$ genes, including those with highest methylation rates, namely *ASCL2, APCDD1* and *PTPRO*. Likewise, *EPHB2* was also silenced at the RNA and protein levels by hypoxia as previously hypothesized (Kaidi et al., 2007). This surprising discovery allowed us to directly link hypoxia to the silencing of WNT/colorectal cancer stem cell genes. This fits well with the fact that hypoxia has been described as an inducer of cell migration (Chaffer and Weinberg, 2011). Likewise our observation of increase EMT related genes (e.g. *SNAIL2* and *ZEB2*) and motility genes (e.g. *HGF* and *SDF1*) in tumors with low WNT ON$^{HR<-1}$ genes support this notion. Thus, it is tempting to speculate that CRC-SCs gain motility at the expense of a transient loss of some stem genes. Markedly, the effects of hypoxia were reversible once oxygen levels were restored. This implies that this molecular switch is reversible and that these “migrating cancer stem cells” can likely revert to their original stem state though an EMT-MET as previously hypothesized (Brabletz et al., 2001; Brabletz et al., 2005).

We tracked hypoxic cells in mice to invasion fronts of orthotopic models and CRC tumors. Hypoxia induced genes LAMC2, CXCR4 and GLUT1 marked budding cells at the invasion fronts, while the WNT/CRC-SC gene EPHB2 (Guo et al., 2006; Karamitopoulou et al., 2010) and pan epithelial marker EPCAM (Hostettler et al., 2010) was silenced in invading cells despite strong nuclear beta-catenin accumulation (Brabletz et al., 2001) as previously reported. Importantly, EPHB2 continued to be expressed in glandular structures present in the tumor bulk (Merlos-Suarez et al., 2011). Hypoxia response signatures from either LS174T cancer cells or primary CRC-SCs were remarkably good predictors of disease relapse and identified those patients with the lowest WNT ON$^{HR<-1}$ genes. Collectively, it is clear that hypoxic conditions in CRC lead to a decrease in a subset on WNT target genes, including *EPHB2*, and a more aggressive phenotype. Interestingly, we also observed heterogeneous differentiation states in budding CRC cells based on KRT20 staining. Our observations raise fundamental questions such as: 1) Does a hierarchy exist in budding tumor cells? 2) What extent of stem-like features might hypoxia/EMT provide to more differentiated cells (i.e. KTR20-high)? 3) Are the KRT20-negative CRC-SCs responsible for metastatic spread? 4) Do these apparently divergent differentiation states provide an essential microenvironment for CRC-SC survival? 5) Are the budding cells at the invasion front the “cells of origin” of liver and/or lung metastasis in CRC?

Clearly targeting hypoxia in CRC should be a primary goal in the field considering our work and the strong correlation with poor prognosis and disease relapse for a variety of hypoxia target genes in CRC (Baba et al., 2010; Furudoi et al., 2001; Koukourakis et al., 2006). Great promise comes from work performed on carbonic anhydrase 9 (CA9) one of the downstream targets of hypoxia that is responsible for counteracting acidosis through regulation of intra- and extra-cellular pH under hypoxia. The silencing of *CA9* using inducible shRNA constructs resulted in drastic reduction in tumor volume in hypoxic xenografts from LS174T (Chiche et al., 2009). In a very recent study reduction of CA9 in CRC cell line xenografts enhances anti-VEGF therapy (McIntyre et al., 2012). In addition, small molecule inhibitors that selectively target CA9 have been shown to inhibit tumor growth and metastasis in breast cancer models (Lou et al., 2011). This study exemplifies the use of targeting downstream effectors of HIF1-alpha and supports the utility of targeting hypoxic cells in tumors. CA9 expression has been associated with poor prognosis in a multitude of cancers (McDonald et al., 2012) including colorectal cancer (Korkeila et al., 2009; Saarnio et al., 1998). In fact, based on the presented data such anti-hypoxic therapeutic strategies may prove beneficial to prevent aggressive CRC tumors and metastasis. Anti-hypoxic strategies and an up to date list of anti-hypoxia drugs in pre-clinical and clinical trials have recently been reviewed (Wilson and Hay, 2011).

In the era of personalized medicine a major obstacle is patient stratification (i.e. identifying those who would best benefit from therapy). The use of microarray signatures could most certainly aid in stratification of patients based on their expression levels of hypoxia related genes (e.g. our described HYP-RSs) or silencing of WNT ON genes associated to poor prognosis (e.g. our describes WNT ON$^{HR<-1}$ signature), although this approach would require surgical intervention. Alternatively, live non-invasive imaging with fluorescent labelled CA9 antibodies and selective CA9 inhibitors, such as HS680, have recently been successfully achieved (Bao et al., 2012; Tafreshi et al., 2012). Transferring such non-invasive diagnostic technologies to clinic is actively being pursued as a result of the clinical implication of tumor hypoxia.

# REFERENCES


Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America *100*, 3983-3988.

Alazzouzi, H., Davalos, V., Kokko, A., Domingo, E., Woerner, S.M., Wilson, A.J., Konrad, L., Laiho, P., Espin, E., Armengol, M.*, et al.* (2005). Mechanisms of inactivation of the receptor tyrosine kinase EPHB2 in colorectal tumors. Cancer research *65*, 10170-10173.

Alexandrovich, A., Qureishi, A., Coudert, A.E., Zhang, L., Grigoriadis, A.E., Shah, A.M., Brewer, A.C., and Pizzey, J.A. (2008). A role for GATA-6 in vertebrate chondrogenesis. Developmental biology *314*, 457-470.

Anttonen, M., Unkila-Kallio, L., Leminen, A., Butzow, R., and Heikinheimo, M. (2005). High GATA-4 expression associates with aggressive behavior, whereas low anti-Mullerian hormone expression associates with growth potential of ovarian granulosa cell tumors. The Journal of clinical endocrinology and metabolism *90*, 6529-6535.

Arnoux, V., Nassour, M., L'Helgoualc'h, A., Hipskind, R.A., and Savagner, P. (2008). Erk5 controls Slug expression and keratinocyte activation during wound healing. Molecular biology of the cell *19*, 4738-4749.

Assou, S., Le Carrour, T., Tondeur, S., Strom, S., Gabelle, A., Marty, S., Nadal, L., Pantesco, V., Reme, T., Hugnot, J.P.*, et al.* (2007). A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem cells *25*, 961-973.

Aubin, J., Davy, A., and Soriano, P. (2004). In vivo convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. Genes & development *18*, 1482-1494.

Auclair, B.A., Benoit, Y.D., Rivard, N., Mishina, Y., and Perreault, N. (2007). Bone morphogenetic protein signaling is essential for terminal differentiation of the intestinal secretory cell lineage. Gastroenterology *133*, 887-896.

Baba, Y., Nosho, K., Shima, K., Irahara, N., Chan, A.T., Meyerhardt, J.A., Chung, D.C., Giovannucci, E.L., Fuchs, C.S., and Ogino, S. (2010). HIF1A overexpression is associated with poor prognosis in a cohort of 731 colorectal cancers. The American journal of pathology *176*, 2292-2301.

Bailey, J.M., Singh, P.K., and Hollingsworth, M.A. (2007). Cancer metastasis facilitated by developmental pathways: Sonic hedgehog, Notch, and bone morphogenic proteins. Journal of cellular biochemistry *102*, 829-839.

Bao, B., Groves, K., Zhang, J., Handy, E., Kennedy, P., Cuneo, G., Supuran, C.T., Yared, W., Rajopadhye, M., and Peterson, J.D. (2012). In vivo imaging and quantification of carbonic anhydrase IX expression as an endogenous biomarker of tumor hypoxia. PloS one *7*, e50860.

Barker, N., Ridgway, R.A., van Es, J.H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. Nature *457*, 608-611.

Barker, N., van Es, J.H., Jaks, V., Kasper, M., Snippert, H., Toftgard, R., and Clevers, H. (2008). Very long-term self-renewal of small intestine, colon, and hair follicles from cycling Lgr5+ve stem cells. Cold Spring Harbor symposia on quantitative biology *73*, 351-356.

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*., et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

Barker, N., van Oudenaarden, A., and Clevers, H. (2012). Identifying the stem cell of the intestinal crypt: strategies and pitfalls. Cell stem cell *11*, 452-460.

Barrallo-Gimeno, A., and Nieto, M.A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development *132*, 3151-3161.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res *35*, D760-765.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nature genetics *37*, 382-390.

Bataille, F., Rohrmeier, C., Bates, R., Weber, A., Rieder, F., Brenmoehl, J., Strauch, U., Farkas, S., Furst, A., Hofstadter, F*., et al.* (2008). Evidence for a role of epithelial mesenchymal transition during pathogenesis of fistulae in Crohn's disease. Inflammatory bowel diseases *14*, 1514-1527.

Batlle, E., Bacani, J., Begthel, H., Jonkheer, S., Gregorieff, A., van de Born, M., Malats, N., Sancho, E., Boon, E., Pawson, T*., et al.* (2005). EphB receptor activity suppresses colorectal cancer progression. Nature *435*, 1126-1130.

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*., et al.* (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell *111*, 251-263.

Batlle, E., Sancho, E., Franci, C., Dominguez, D., Monfar, M., Baulida, J., and Garcia De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nature cell biology *2*, 84-89.

Batts, L.E., Polk, D.B., Dubois, R.N., and Kulessa, H. (2006). Bmp signaling is required for intestinal growth and morphogenesis. Developmental dynamics : an official publication of the American Association of Anatomists *235*, 1563-1570.

Beck, S.E., Jung, B.H., Fiorino, A., Gomez, J., Rosario, E.D., Cabrera, B.L., Huang, S.C., Chow, J.Y., and Carethers, J.M. (2006). Bone morphogenetic protein signaling and growth suppression in colon cancer. American journal of physiology Gastrointestinal and liver physiology *291*, G135-145.

Belaguli, N.S., Aftab, M., Rigi, M., Zhang, M., Albo, D., and Berger, D.H. (2010). GATA6 promotes colon cancer cell invasion by regulating urokinase plasminogen activator gene expression. Neoplasia *12*, 856-865.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nature genetics *40*, 499-507.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics *29*, 1165–1188.

Benson, A.B., 3rd, Schrag, D., Somerfield, M.R., Cohen, A.M., Figueredo, A.T., Flynn, P.J., Krzyzanowska, M.K., Maroun, J., McAllister, P., Van Cutsem, E.*, et al.* (2004). American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *22*, 3408-3419.

Beppu, H., Mwizerwa, O.N., Beppu, Y., Dattwyler, M.P., Lauwers, G.Y., Bloch, K.D., and Goldstein, A.M. (2008). Stromal inactivation of BMPRII leads to colorectal epithelial overgrowth and polyp formation. Oncogene *27*, 1063-1070.

Beuling, E., Aronson, B.E., Tran, L.M., Stapleton, K.A., ter Horst, E.N., Vissers, L.A., Verzi, M.P., and Krasinski, S.D. (2012). GATA6 is required for proliferation, migration, secretory cell maturation, and gene expression in the mature mouse colon. Molecular and cellular biology *32*, 3392-3402.

Beuling, E., Baffour-Awuah, N.Y., Stapleton, K.A., Aronson, B.E., Noah, T.K., Shroyer, N.F., Duncan, S.A., Fleet, J.C., and Krasinski, S.D. (2011). GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. Gastroenterology *140*, 1219-1229 e1211-1212.

Beuling, E., Kerkhof, I.M., Nicksa, G.A., Giuffrida, M.J., Haywood, J., aan de Kerk, D.J., Piaseckyj, C.M., Pu, W.T., Buchmiller, T.L., Dawson, P.A.*, et al.* (2010). Conditional Gata4 deletion in mice induces bile acid absorption in the proximal small intestine. Gut *59*, 888-895.

Bhanot, P., Brink, M., Samos, C.H., Hsieh, J.C., Wang, Y., Macke, J.P., Andrew, D., Nathans, J., and Nusse, R. (1996). A new member of the frizzled family from Drosophila functions as a Wingless receptor. Nature *382*, 225-230.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., and Gunderson, K.L. (2009). Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics *1*, 177-200.

Bolos, V., Peinado, H., Perez-Moreno, M.A., Fraga, M.F., Esteller, M., and Cano, A. (2003). The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. Journal of cell science *116*, 499-511.

Bond, J.H. (2000). Polyp guideline: diagnosis, treatment, and surveillance for patients with colorectal polyps. Practice Parameters Committee of the American College of Gastroenterology. The American journal of gastroenterology *95*, 3053-3063.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature medicine *3*, 730-737.

Bosse, T., Piaseckyj, C.M., Burghard, E., Fialkovich, J.J., Rajagopal, S., Pu, W.T., and Krasinski, S.D. (2006). Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. Molecular and cellular biology *26*, 9060-9070.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G*., et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Brabletz, T., Jung, A., Reu, S., Porzner, M., Hlubek, F., Kunz-Schughart, L.A., Knuechel, R., and Kirchner, T. (2001). Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. Proceedings of the National Academy of Sciences of the United States of America *98*, 10356-10361.

Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., and Kirchner, T. (2005). Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nature reviews Cancer *5*, 744-749.

Brenner, H., Chang-Claude, J., Seiler, C.M., Sturmer, T., and Hoffmeister, M. (2007). Potential for colorectal cancer prevention of sigmoidoscopy versus colonoscopy: population-based case control study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology *16*, 494-499.

Broderick, P., Carvajal-Carmona, L., Pittman, A.M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S*., et al.* (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nature genetics *39*, 1315-1317.

Brosens, L.A., Langeveld, D., van Hattem, W.A., Giardiello, F.M., and Offerhaus, G.J. (2011). Juvenile polyposis syndrome. World journal of gastroenterology : WJG *17*, 4839-4844.

Brosens, L.A., van Hattem, A., Hylind, L.M., Iacobuzio-Donahue, C., Romans, K.E., Axilbund, J., Cruz-Correa, M., Tersmette, A.C., Offerhaus, G.J., and Giardiello, F.M. (2007). Risk of colorectal cancer in juvenile polyposis. Gut *56*, 965-967.

Buczacki, S.J., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature *495*, 65-69.

Burt, R.W. (2000). Colon cancer screening. Gastroenterology *119*, 837-853.
Butterworth, A.S., Higgins, J.P., and Pharoah, P. (2006). Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. European journal of cancer *42*, 216-227.

Cai, J., Pardali, E., Sanchez-Duffhues, G., and ten Dijke, P. (2012). BMP signaling in vascular diseases. FEBS letters *586*, 1993-2002.

Cai, K.Q., Caslini, C., Capo-chichi, C.D., Slater, C., Smith, E.R., Wu, H., Klein-Szanto, A.J., Godwin, A.K., and Xu, X.X. (2009). Loss of GATA4 and GATA6 expression specifies ovarian cancer histological subtypes and precedes neoplastic transformation of ovarian surface epithelia. PloS one *4*, e6454.

Cairns, R.A., Harris, I.S., and Mak, T.W. (2011). Regulation of cancer cell metabolism. Nature reviews Cancer *11*, 85-95.

Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K., and Kaufman, P.D. (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. PloS one *4*, e6529.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The Embryonic development of Drosophila melanogaster (Berlin: Springer-Verlag).

Campos-Ortega, J.A., and Hartenstein, V. (1997). The Embryonic Development of *Drosophila melanogaster*, 2nd edn (Springer-Verlag).

Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330-337.

Cano, A., Perez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., and Nieto, M.A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nature cell biology *2*, 76-83.

Cao, C., Liu, Y., and Lehmann, M. (2007). Fork head controls the timing and tissue selectivity of steroid-induced developmental cell death. J Cell Biol *176*, 843-852.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H*., et al.* (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318-325.

Carver, E.A., Jiang, R., Lan, Y., Oram, K.F., and Gridley, T. (2001). The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. Molecular and cellular biology *21*, 8184-8188.

CGARN (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061-1068.

Chaffer, C.L., and Weinberg, R.A. (2011). A perspective on cancer cell metastasis. Science *331*, 1559-1564.

Chang, H., Brown, C.W., and Matzuk, M.M. (2002). Genetic analysis of the mammalian transforming growth factor-beta superfamily. Endocrine reviews *23*, 787-823.

Chang, Q., Jurisica, I., Do, T., and Hedley, D.W. (2011). Hypoxia predicts aggressive growth and spontaneous metastasis formation from orthotopically grown primary xenografts of human pancreatic cancer. Cancer research *71*, 3110-3120.

Chaudary, N., and Hill, R.P. (2009). Increased expression of metastasis-related genes in hypoxic cells sorted from cervical and lymph nodal xenograft tumors. Lab Invest *89*, 587-596.

Chen, Y., Bates, D.L., Dey, R., Chen, P.H., Machado, A.C., Laird-Offringa, I.A., Rohs, R., and Chen, L. (2012). DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. Cell reports *2*, 1197-1206.

Cheng, H., and Leblond, C.P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. III. Entero-endocrine cells. The American journal of anatomy *141*, 503-519.

Chiche, J., Ilc, K., Laferriere, J., Trottier, E., Dayan, F., Mazure, N.M., Brahimi-Horn, M.C., and Pouyssegur, J. (2009). Hypoxia-inducible carbonic anhydrase IX and XII promote tumor cell growth by counteracting acidosis through the regulation of the intracellular pH. Cancer research *69*, 358-368.

Chikina, M.D., and Troyanskaya, O.G. (2012). An effective statistical evaluation of ChIPseq dataset similarity. Bioinformatics *28*, 607-613.

Chiu, S.T., Chang, K.J., Ting, C.H., Shen, H.C., Li, H., and Hsieh, F.J. (2009). Over-expression of EphB3 enhances cell-cell contacts and suppresses tumor growth in HT-29 human colon cancer cells. Carcinogenesis *30*, 1475-1486.

Choi, S.S., and Diehl, A.M. (2009). Epithelial-to-mesenchymal transitions in the liver. Hepatology *50*, 2007-2013.

Clarke, M.F., Dick, J.E., Dirks, P.B., Eaves, C.J., Jamieson, C.H., Jones, D.L., Visvader, J., Weissman, I.L., and Wahl, G.M. (2006). Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer research *66*, 9339-9344.

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469-480.

Clevers, H., and Batlle, E. (2006). EphB/EphrinB receptors and Wnt signaling in colorectal cancer. Cancer research *66*, 2-5.

Clevers, H., and Batlle, E. (2013). SnapShot: the intestinal crypt. Cell *152*, 1198-1198 e1192.

Collins, A.T., Berry, P.A., Hyde, C., Stower, M.J., and Maitland, N.J. (2005). Prospective identification of tumorigenic prostate cancer stem cells. Cancer research *65*, 10946-10951.

Colnot, S., Decaens, T., Niwa-Kawakita, M., Godard, C., Hamard, G., Kahn, A., Giovannini, M., and Perret, C. (2004). Liver-targeted disruption of Apc in mice activates beta-catenin signaling and leads to hepatocellular carcinomas. Proceedings of the National Academy of Sciences of the United States of America *101*, 17216-17221.

Cortina, C., Palomo-Ponce, S., Iglesias, M., Fernandez-Masip, J.L., Vivancos, A., Whissell, G., Huma, M., Peiro, N., Gallego, L., Jonkheer, S.*, et al.* (2007). EphB-ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells. Nature genetics *39*, 1376-1383.

Cox, D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society Series B *34*, 187-220.

Cross, E.E., Thomason, R.T., Martinez, M., Hopkins, C.R., Hong, C.C., and Bader, D.M. (2011). Application of small organic molecules reveals cooperative TGFbeta and BMP regulation of mesothelial cell behaviors. ACS chemical biology *6*, 952-961.

Curtin, J.C. (2013). Novel drug discovery opportunities for colorectal cancer. Expert opinion on drug discovery *8*, 1153-1164.

Dalerba, P., Cho, R.W., and Clarke, M.F. (2007a). Cancer stem cells: models and concepts. Annual review of medicine *58*, 267-284.

Dalerba, P., Dylla, S.J., Park, I.K., Liu, R., Wang, X., Cho, R.W., Hoey, T., Gurney, A., Huang, E.H., Simeone, D.M.*, et al.* (2007b). Phenotypic characterization of human colorectal cancer stem cells. Proceedings of the National Academy of Sciences of the United States of America *104*, 10158-10163.

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D.*, et al.* (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol *29*, 1120-1127.

De Biase, P., and Capanna, R. (2005). Clinical applications of BMPs. Injury *36 Suppl 3*, S43-46.

de Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M.*, et al.* (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293-297.

de Sousa, E.M.F., Colak, S., Buikhuisen, J., Koster, J., Cameron, K., de Jong, J.H., Tuynman, J.B., Prasetyanti, P.R., Fessler, E., van den Bergh, S.P.*, et al.* (2011). Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. Cell stem cell *9*, 476-485.

Della Gatta, G., Palomero, T., Perez-Garcia, A., Ambesi-Impiombato, A., Bansal, M., Carpenter, Z.W., De Keersmaecker, K., Sole, X., Xu, L., Paietta, E.*, et al.* (2012). Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. Nature medicine *18*, 436-440.

Demidov, O.N., Timofeev, O., Lwin, H.N., Kek, C., Appella, E., and Bulavin, D.V. (2007). Wip1 phosphatase regulates p53-dependent apoptosis of stem cells and tumorigenesis in the mouse intestine. Cell stem cell *1*, 180-190.

Derner, R., and Anderson, A.C. (2005). The bone morphogenic protein. Clinics in podiatric medicine and surgery *22*, 607-618, vii.

Durand, A., Donahue, B., Peignon, G., Letourneur, F., Cagnard, N., Slomianny, C., Perret, C., Shroyer, N.F., and Romagnolo, B. (2012). Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). Proceedings of the National Academy of Sciences of the United States of America *109*, 8965-8970.

el Marjou, F., Janssen, K.P., Chang, B.H., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Eng, C., and Peacocke, M. (1998). PTEN and inherited hamartoma-cancer syndromes. Nature genetics *19*, 223.

Erler, J.T., Bennewith, K.L., Nicolau, M., Dornhofer, N., Kong, C., Le, Q.T., Chi, J.T., Jeffrey, S.S., and Giaccia, A.J. (2006). Lysyl oxidase is essential for hypoxia-induced metastasis. Nature *440*, 1222-1226.

Erler, J.T., Cawthorne, C.J., Williams, K.J., Koritzinsky, M., Wouters, B.G., Wilson, C., Miller, C., Demonacos, C., Stratford, I.J., and Dive, C. (2004). Hypoxia-mediated down-regulation of Bid and Bax in tumors occurs via hypoxia-inducible factor 1-dependent and -independent mechanisms and contributes to drug resistance. Molecular and cellular biology *24*, 2875-2889.

Farrall, A.L., Riemer, P., Leushacke, M., Sreekumar, A., Grimm, C., Herrmann, B.G., and Morkel, M. (2012). Wnt and BMP signals control intestinal adenoma cell fates. International journal of cancer Journal international du cancer *131*, 2242-2252.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759-767.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proceedings of the National Academy of Sciences of the United States of America *93*, 10887-10890.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochemical and biophysical research communications *237*, 752-757.

Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes & development *17*, 1115-1129.

Fidler, I.J., and Poste, G. (2008). The "seed and soil" hypothesis revisited. The lancet oncology *9*, 808.

Frede, S., Freitag, P., Otto, T., Heilmaier, C., and Fandrey, J. (2005). The proinflammatory cytokine interleukin 1beta and hypoxia cooperatively induce the expression of adrenomedullin in ovarian carcinoma cells through hypoxia inducible factor 1 activation. Cancer research *65*, 4690-4697.

Freeman, T.J., Smith, J.J., Chen, X., Washington, M.K., Roland, J.T., Means, A.L., Eschrich, S.A., Yeatman, T.J., Deane, N.G., and Beauchamp, R.D. (2012). Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. Gastroenterology *142*, 562-571 e562.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nature reviews Molecular cell biology *10*, 445-457.

Fu, B., Luo, M., Lakkur, S., Lucito, R., and Iacobuzio-Donahue, C.A. (2008). Frequent genomic copy number gain and overexpression of GATA-6 in pancreatic carcinoma. Cancer Biol Ther *7*, 1593-1601.

Fu, T., Li, P., Wang, H., He, Y., Luo, D., Zhang, A., Tong, W., Zhang, L., Liu, B., and Hu, C. (2009). c-Rel is a transcriptional repressor of EPHB2 in colorectal cancer. The Journal of pathology *219*, 103-113.

Furudoi, A., Tanaka, S., Haruma, K., Yoshihara, M., Sumii, K., Kajiyama, G., and Shimamoto, F. (2001). Clinical significance of human erythrocyte glucose transporter 1 expression at the deepest invasive site of advanced colorectal carcinoma. Oncology *60*, 162-169.

Gao, X., Sedgwick, T., Shi, Y.B., and Evans, T. (1998). Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. Molecular and cellular biology *18*, 2901-2911.

Gentile, A., Trusolino, L., and Comoglio, P.M. (2008). The Met tyrosine kinase receptor in development and cancer. Cancer metastasis reviews *27*, 85-94.

Gerbe, F., Brulin, B., Makrini, L., Legraverend, C., and Jay, P. (2009). DCAMKL-1 expression identifies Tuft cells rather than stem cells in the adult mouse intestinal epithelium. Gastroenterology *137*, 2179-2180; author reply 2180-2171.

Gerbe, F., Legraverend, C., and Jay, P. (2012). The intestinal epithelium tuft cells: specification and function. Cellular and molecular life sciences : CMLS *69*, 2907-2917.

Giannakis, M., Stappenbeck, T.S., Mills, J.C., Leip, D.G., Lovett, M., Clifton, S.W., Ippolito, J.E., Glasscock, J.I., Arumugam, M., Brent, M.R*., et al.* (2006). Molecular properties of adult mouse gastric and intestinal epithelial progenitors in their niches. The Journal of biological chemistry *281*, 11292-11300.

Gillis, W.Q., Bowerman, B.A., and Schneider, S.Q. (2008). The evolution of protostome GATA factors: molecular phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. BMC Evol Biol *8*, 112.

Gonzalez-Roca, E., Garcia-Albeniz, X., Rodriguez-Mulero, S., Gomis, R.R., Kornacker, K., and Auer, H. Accurate expression profiling of very small cell populations. PloS one *5*, e14418.

Gonzalez-Roca, E., Garcia-Albeniz, X., Rodriguez-Mulero, S., Gomis, R.R., Kornacker, K., and Auer, H. (2010). Accurate expression profiling of very small cell populations. PloS one *5*, e14418.

Gort, E.H., Groot, A.J., van der Wall, E., van Diest, P.J., and Vooijs, M.A. (2008). Hypoxic regulation of metastasis via hypoxia-inducible factors. Current molecular medicine *8*, 60-67.

Grawe, F., Wodarz, A., Lee, B., Knust, E., and Skaer, H. (1996). The Drosophila genes crumbs and stardust are involved in the biogenesis of adherens junctions. Development *122*, 951-959.

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.

Gregory, P.A., Bracken, C.P., Bert, A.G., and Goodall, G.J. (2008). MicroRNAs as regulators of epithelial-mesenchymal transition. Cell cycle *7*, 3112-3118.

Grooteclaes, M.L., and Frisch, S.M. (2000). Evidence for a function of CtBP in epithelial gene regulation and anoikis. Oncogene *19*, 3823-3828.

Groppe, J., Greenwald, J., Wiater, E., Rodriguez-Leon, J., Economides, A.N., Kwiatkowski, W., Affolter, M., Vale, W.W., Izpisua Belmonte, J.C., and Choe, S. (2002). Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. Nature *420*, 636-642.

Groppe, J., Greenwald, J., Wiater, E., Rodriguez-Leon, J., Economides, A.N., Kwiatkowski, W., Baban, K., Affolter, M., Vale, W.W., Izpisua Belmonte, J.C*., et al.* (2003). Structural basis of BMP signaling inhibition by Noggin, a novel twelve-membered cystine knot protein. The Journal of bone and joint surgery American volume *85-A Suppl 3*, 52-58.

Guarino, M., Rubino, B., and Ballabio, G. (2007). The role of epithelial-mesenchymal transition in cancer pathology. Pathology *39*, 305-318.

Guo, D.L., Zhang, J., Yuen, S.T., Tsui, W.Y., Chan, A.S., Ho, C., Ji, J., Leung, S.Y., and Chen, X. (2006). Reduced expression of EphB2 that parallels invasion and metastasis in colorectal tumours. Carcinogenesis *27*, 454-464.

Guzy, R.D., Hoyos, B., Robin, E., Chen, H., Liu, L., Mansfield, K.D., Simon, M.C., Hammerling, U., and Schumacker, P.T. (2005). Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. Cell metabolism *1*, 401-408.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hanawa, H., Kelly, P.F., Nathwani, A.C., Persons, D.A., Vandergriff, J.A., Hargrove, P., Vanin, E.F., and Nienhuis, A.W. (2002). Comparison of various envelope proteins for their ability to pseudotype lentiviral vectors and transduce primitive hematopoietic cells from human blood. Mol Ther *5*, 242-251.

Hao, J., Ho, J.N., Lewis, J.A., Karim, K.A., Daniels, R.N., Gentry, P.R., Hopkins, C.R., Lindsley, C.W., and Hong, C.C. (2010). In vivo structure-activity relationship study of dorsomorphin analogues identifies selective VEGF and BMP inhibitors. ACS chemical biology *5*, 245-253.

Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M., and Taketo, M.M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. The EMBO journal *18*, 5931-5942.

Haramis, A.P., Begthel, H., van den Born, M., van Es, J., Jonkheer, S., Offerhaus, G.J., and Clevers, H. (2004). De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. Science *303*, 1684-1686.

Hardwick, J.C., Kodach, L.L., Offerhaus, G.J., and van den Brink, G.R. (2008). Bone morphogenetic protein signalling in colorectal cancer. Nature reviews Cancer *8*, 806-812.

Hardwick, J.C., Van Den Brink, G.R., Bleuming, S.A., Ballester, I., Van Den Brande, J.M., Keller, J.J., Offerhaus, G.J., Van Deventer, S.J., and Peppelenbosch, M.P. (2004). Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon. Gastroenterology *126*, 111-121.

Hartenstein V. (1993). Atlas of Drosophila development (New York: Cold Spring Harbor Laboratory Press ).

Haveri, H., Westerholm-Ormio, M., Lindfors, K., Maki, M., Savilahti, E., Andersson, L.C., and Heikinheimo, M. (2008). Transcription factors GATA-4 and GATA-6 in normal and neoplastic human gastrointestinal mucosa. BMC gastroenterology *8*, 9.

Hay, E.D. (1995). An overview of epithelio-mesenchymal transformation. Acta Anat (Basel) *154*, 8-20.

He, X.C., Zhang, J., Tong, W.G., Tawfik, O., Ross, J., Scoville, D.H., Tian, Q., Zeng, X., He, X., Wiedemann, L.M*., et al.* (2004). BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. Nature genetics *36*, 1117-1121.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W*., et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Hellebrekers, D.M., Lentjes, M.H., van den Bosch, S.M., Melotte, V., Wouters, K.A., Daenen, K.L., Smits, K.M., Akiyama, Y., Yuasa, Y., Sanduleanu, S*., et al.* (2009). GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *15*, 3990-3997.

Hewitson, K.S., McNeill, L.A., Riordan, M.V., Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Bhattacharya, S., Gleadle, J.M*., et al.* (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. The Journal of biological chemistry *277*, 26351-26355.

Hill, R.P., Marie-Egyptienne, D.T., and Hedley, D.W. (2009). Cancer stem cells, hypoxia and metastasis. Seminars in radiation oncology *19*, 106-111.

Hochberg, Y.B.a.Y. (1995). Controlling the false discovery rate in behavior genetics research. Journal of the Royal Statistical Society B *57*, 289–300.

Hockel, M., and Vaupel, P. (2001). Tumor hypoxia: definitions and current clinical, biologic, and molecular aspects. Journal of the National Cancer Institute *93*, 266-276.

Horst, D., Kriegl, L., Engel, J., Jung, A., and Kirchner, T. (2009a). CD133 and nuclear beta-catenin: the marker combination to detect high risk cases of low stage colorectal cancer. European journal of cancer *45*, 2034-2040.

Horst, D., Kriegl, L., Engel, J., Kirchner, T., and Jung, A. (2009b). Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. Cancer investigation *27*, 844-850.

Horst, D., Scheel, S.K., Liebmann, S., Neumann, J., Maatz, S., Kirchner, T., and Jung, A. (2009c). The cancer stem cell marker CD133 has high prognostic impact but unknown functional relevance for the metastasis of human colon cancer. The Journal of pathology *219*, 427-434.

Hostettler, L., Zlobec, I., Terracciano, L., and Lugli, A. (2010). ABCG5-positivity in tumor buds is an indicator of poor prognosis in node-negative colorectal cancer patients. World journal of gastroenterology : WJG *16*, 732-739.

Houlston, R.S., Cheadle, J., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Spain, S.L., Broderick, P., Domingo, E., Farrington, S*., et al.* (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nature genetics *42*, 973-977.

Howe, J.R., Bair, J.L., Sayed, M.G., Anderson, M.E., Mitros, F.A., Petersen, G.M., Velculescu, V.E., Traverso, G., and Vogelstein, B. (2001). Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nature genetics *28*, 184-187.

Huang, S.C., Chen, C.R., Lavine, J.E., Taylor, S.F., Newbury, R.O., Pham, T.T., Ricciardiello, L., and Carethers, J.M. (2000). Genetic heterogeneity in familial juvenile polyposis. Cancer research *60*, 6882-6885.

Huang, W.S., Chin, C.C., Chen, C.N., Kuo, Y.H., Chen, T.C., Yu, H.R., Tung, S.Y., Shen, C.H., Hsieh, Y.Y., Guo, S.E*., et al.* (2012). Stromal cell-derived factor-1/CXC receptor 4 and beta1 integrin interaction regulates urokinase-type plasminogen activator expression in human colorectal cancer cells. Journal of cellular physiology *227*, 1114-1122.

Huber, M.A., Kraut, N., and Beug, H. (2005). Molecular requirements for epithelial-mesenchymal transition during tumor progression. Curr Opin Cell Biol *17*, 548-558.

Hyvonen, M. (2003). CHRD, a novel domain in the BMP inhibitor chordin, is also found in microbial proteins. Trends in biochemical sciences *28*, 470-473.

Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M., and Miyazono, K. (1997). Smad6 inhibits signalling by the TGF-beta superfamily. Nature *389*, 622-626.

Imperiale, T.F., Wagner, D.R., Lin, C.Y., Larkin, G.N., Rogge, J.D., and Ransohoff, D.F. (2000). Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. The New England journal of medicine *343*, 169-174.

Ireland, H., Houghton, C., Howard, L., and Winton, D.J. (2005). Cellular inheritance of a Cre-activated reporter gene to determine Paneth cell longevity in the murine small intestine. Developmental dynamics : an official publication of the American Association of Anatomists *233*, 1332-1336.

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.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2012). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nature cell biology *14*, 106-114.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G., Jr. (2001). HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science *292*, 464-468.

Iwamoto, M., Hoffenberg, E.J., Carethers, J.M., Doctolero, R., Tajima, A., Sugano, K., Franklin, W.A., and Ahnen, D.J. (2005). Nuclear accumulation of beta-catenin occurs commonly in the epithelial cells of juvenile polyps. Pediatric research *57*, 4-9; discussion 1-3.

Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J.*, et al.* (2001). Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science *292*, 468-472.

Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P.R., Davis, H., Kaur, K., Heinimann, K., Howarth, K.*, et al.* (2012). Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nature genetics *44*, 699-703.

Jain, R.K. (2005). Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science *307*, 58-62.

Jasperson, K.W., Tuohy, T.M., Neklason, D.W., and Burt, R.W. (2010). Hereditary and familial colon cancer. Gastroenterology *138*, 2044-2058.

Jass, J.R., Barker, M., Fraser, L., Walsh, M.D., Whitehall, V.L., Gabrielli, B., Young, J., and Leggett, B.A. (2003). APC mutation and tumour budding in colorectal cancer. Journal of clinical pathology *56*, 69-73.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. CA: a cancer journal for clinicians *61*, 69-90.

Jubb, A.M., Zhong, F., Bheddah, S., Grabsch, H.I., Frantz, G.D., Mueller, W., Kavi, V., Quirke, P., Polakis, P., and Koeppen, H. (2005). EphB2 is a prognostic factor in colorectal cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *11*, 5181-5187.

Jung, P., Sato, T., Merlos-Suarez, A., Barriga, F.M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M.A., Sancho, E*., et al.* (2011). Isolation and in vitro expansion of human colonic stem cells. Nature medicine *17*, 1225-1227.

Kaelin, W.G., Jr., and Ratcliffe, P.J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Molecular cell *30*, 393-402.

Kahi, C.J., Imperiale, T.F., Juliar, B.E., and Rex, D.K. (2009). Effect of screening colonoscopy on colorectal cancer incidence and mortality. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association *7*, 770-775; quiz 711.

Kahlert, C., Lahes, S., Radhakrishnan, P., Dutta, S., Mogler, C., Herpel, E., Brand, K., Steinert, G., Schneider, M., Mollenhauer, M*., et al.* (2011). Overexpression of ZEB2 at the invasion front of colorectal cancer is an independent prognostic marker and regulates tumor invasion in vitro. Clinical cancer research : an official journal of the American Association for Cancer Research *17*, 7654-7663.

Kaidi, A., Moorghen, M., Williams, A.C., and Paraskeva, C. (2007). Is the downregulation of EphB2 receptor expression during colorectal tumorigenesis due to hypoxia? Gut *56*, 1637-1638.

Kalluri, R., and Neilson, E.G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. The Journal of clinical investigation *112*, 1776-1784.

Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. The Journal of clinical investigation *119*, 1420-1428.

Kamnasaran, D., Qian, B., Hawkins, C., Stanford, W.L., and Guha, A. (2007). GATA6 is an astrocytoma tumor suppressor gene identified by gene trapping of mouse glioma model. Proceedings of the National Academy of Sciences of the United States of America *104*, 8053-8058.

Karagiannis, G.S., Berk, A., Dimitromanolakis, A., and Diamandis, E.P. (2013). Enrichment map profiling of the cancer invasion front suggests regulation of colorectal cancer progression by the bone morphogenetic protein antagonist, gremlin-1. Molecular oncology *7*, 826-839.

Karam, S.M. (1999). Lineage commitment and maturation of epithelial cells in the gut. Frontiers in bioscience : a journal and virtual library *4*, D286-298.

Karamitopoulou, E., Lugli, A., Panayiotides, I., Karakitsos, P., Peros, G., Rallis, G., Patsouris, E.S., Terracciano, L., and Zlobec, I. (2010). Systematic assessment of protein phenotypes characterizing high-grade tumour budding in mismatch repair-proficient colorectal cancer. Histopathology *57*, 233-243.

Karbanova, J., Missol-Kolka, E., Fonseca, A.V., Lorra, C., Janich, P., Hollerova, H., Jaszai, J., Ehrmann, J., Kolar, Z., Liebers, C*., et al.* (2008). The stem cell marker CD133 (Prominin-1) is expressed in various human glandular epithelia. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *56*, 977-993.

Kazama, S., Watanabe, T., Ajioka, Y., Kanazawa, T., and Nagawa, H. (2006). Tumour budding at the deepest invasive margin correlates with lymph node metastasis in submucosal colorectal cancer detected by anticytokeratin antibody CAM5.2. British journal of cancer *94*, 293-298.

Keith, B., Johnson, R.S., and Simon, M.C. (2012). HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. Nature reviews Cancer *12*, 9-22.

Kelley, C., Blumberg, H., Zon, L.I., and Evans, T. (1993). GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. Development *118*, 817-827.

Kelloff, G.J., Schilsky, R.L., Alberts, D.S., Day, R.W., Guyton, K.Z., Pearce, H.L., Peck, J.C., Phillips, R., and Sigman, C.C. (2004). Colorectal adenomas: a prototype for the use of surrogate end points in the development of cancer prevention drugs. Clinical cancer research : an official journal of the American Association for Cancer Research *10*, 3908-3918.

Kemper, K., Prasetyanti, P.R., De Lau, W., Rodermond, H., Clevers, H., and Medema, J.P. (2012). Monoclonal antibodies against Lgr5 identify human colorectal cancer stem cells. Stem cells *30*, 2378-2386.

Kevans, D., Wang, L.M., Sheahan, K., Hyland, J., O'Donoghue, D., Mulcahy, H., and O'Sullivan, J. (2011). Epithelial-mesenchymal transition (EMT) protein expression in a cohort of stage II colorectal cancer patients with characterized tumor budding and mismatch repair protein status. International journal of surgical pathology *19*, 751-760.

Kim, B.G., Li, C., Qiao, W., Mamura, M., Kasprzak, B., Anver, M., Wolfraim, L., Hong, S., Mushinski, E., Potter, M*., et al.* (2006). Smad4 signalling in T cells is required for suppression of gastrointestinal cancer. Nature *441*, 1015-1019.

Kim, J.S., Crooks, H., Dracheva, T., Nishanian, T.G., Singh, B., Jen, J., and Waldman, T. (2002). Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. Cancer research *62*, 2744-2748.

Kim, M., and Choe, S. (2011). BMPs and their clinical potentials. BMB reports *44*, 619-634.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell *87*, 159-170.

Kirkbride, K.C., Townsend, T.A., Bruinsma, M.W., Barnett, J.V., and Blobe, G.C. (2008). Bone morphogenetic proteins signal through the transforming growth factor-beta type III receptor. The Journal of biological chemistry *283*, 7628-7637.

Kodach, L.L., Bleuming, S.A., Musler, A.R., Peppelenbosch, M.P., Hommes, D.W., van den Brink, G.R., van Noesel, C.J., Offerhaus, G.J., and Hardwick, J.C. (2008a). The bone morphogenetic protein pathway is active in human colon adenomas and inactivated in colorectal cancer. Cancer *112*, 300-306.

Kodach, L.L., Wiercinska, E., de Miranda, N.F., Bleuming, S.A., Musler, A.R., Peppelenbosch, M.P., Dekker, E., van den Brink, G.R., van Noesel, C.J., Morreau, H.*, et al.* (2008b). The bone morphogenetic protein pathway is inactivated in the majority of sporadic colorectal cancers. Gastroenterology *134*, 1332-1341.

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. Nature genetics *19*, 379-383.

Korkeila, E., Talvinen, K., Jaakkola, P.M., Minn, H., Syrjanen, K., Sundstrom, J., and Pyrhonen, S. (2009). Expression of carbonic anhydrase IX suggests poor outcome in rectal cancer. British journal of cancer *100*, 874-880.

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. Proceedings of the National Academy of Sciences of the United States of America *104*, 15418-15423.

Koukourakis, M.I., Giatromanolaki, A., Sivridis, E., Gatter, K.C., Harris, A.L., and Tumour Angiogenesis Research, G. (2006). Lactate dehydrogenase 5 expression in operable colorectal cancer: strong association with survival and activated vascular endothelial growth factor pathway--a report of the Tumour Angiogenesis Research Group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *24*, 4301-4308.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. Development *126*, 723-732.

Kretzschmar, K., and Watt, F.M. (2012). Lineage tracing. Cell *148*, 33-45.

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C., and Leiden, J.M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & development *11*, 1048-1060.

Kwei, K.A., Bashyam, M.D., Kao, J., Ratheesh, R., Reddy, E.C., Kim, Y.H., Montgomery, K., Giacomini, C.P., Choi, Y.L., Chatterjee, S.*, et al.* (2008). Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreatobiliary cancer. PLoS genetics *4*, e1000081.

Laforest, B., Andelfinger, G., and Nemer, M. (2011). Loss of Gata5 in mice leads to bicuspid aortic valve. The Journal of clinical investigation *121*, 2876-2887.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M.*, et al.* (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511-515.

Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., Whitelaw, M.L., and Bruick, R.K. (2002). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. Genes & development *16*, 1466-1471.

Langenfeld, E., Hong, C.C., Lanke, G., and Langenfeld, J. (2013). Bone morphogenetic protein type I receptor antagonists decrease growth and induce cell death of lung cancer cell lines. PloS one *8*, e61256.

Lango Allen, H., Flanagan, S.E., Shaw-Smith, C., De Franco, E., Akerman, I., Caswell, R., International Pancreatic Agenesis, C., Ferrer, J., Hattersley, A.T., and Ellard, S. (2012). GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nature genetics *44*, 20-22.

Larrain, J., Bachiller, D., Lu, B., Agius, E., Piccolo, S., and De Robertis, E.M. (2000). BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. Development *127*, 821-830.

Laverriere, A.C., MacNeill, C., Mueller, C., Poelmann, R.E., Burch, J.B., and Evans, T. (1994). GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. The Journal of biological chemistry *269*, 23177-23184.

Lee, N.Y., Kirkbride, K.C., Sheu, R.D., and Blobe, G.C. (2009). The transforming growth factor-beta type III receptor mediates distinct subcellular trafficking and downstream signaling of activin-like kinase (ALK)3 and ALK6 receptors. Molecular biology of the cell *20*, 4362-4370.

Lee, P.J., Jiang, B.H., Chin, B.Y., Iyer, N.V., Alam, J., Semenza, G.L., and Choi, A.M. (1997). Hypoxia-inducible factor-1 mediates transcriptional activation of the heme oxygenase-1 gene in response to hypoxia. The Journal of biological chemistry *272*, 5375-5381.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Levin, T.G., Powell, A.E., Davies, P.S., Silk, A.D., Dismuke, A.D., Anderson, E.C., Swain, J.R., and Wong, M.H. (2010). Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. Gastroenterology *139*, 2072-2082 e2075.

Levine, J.S., and Ahnen, D.J. (2006). Clinical practice. Adenomatous polyps of the colon. The New England journal of medicine *355*, 2551-2557.

Li, C., Heidt, D.G., Dalerba, P., Burant, C.F., Zhang, L., Adsay, V., Wicha, M., Clarke, M.F., and Simeone, D.M. (2007). Identification of pancreatic cancer stem cells. Cancer research *67*, 1030-1037.

Li, V.S., Ng, S.S., Boersema, P.J., Low, T.Y., Karthaus, W.R., Gerlach, J.P., Mohammed, S., Heck, A.J., Maurice, M.M., Mahmoudi, T.*, et al.* (2012). Wnt signaling through inhibition of beta-catenin degradation in an intact Axin1 complex. Cell *149*, 1245-1256.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.*, et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Lin, L., Bass, A.J., Lockwood, W.W., Wang, Z., Silvers, A.L., Thomas, D.G., Chang, A.C., Lin, J., Orringer, M.B., Li, W.*, et al.* (2012). Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. Proceedings of the National Academy of Sciences of the United States of America *109*, 4251-4256.

Lin, X., Huo, Z., Liu, X., Zhang, Y., Li, L., Zhao, H., Yan, B., Liu, Y., Yang, Y., and Chen, Y.H. (2010). A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. Journal of human genetics *55*, 662-667.

Lindholm, P.M., Soini, Y., Myllarniemi, M., Knuutila, S., Heikinheimo, M., Kinnula, V.L., and Salmenkivi, K. (2009). Expression of GATA-6 transcription factor in pleural malignant mesothelioma and metastatic pulmonary adenocarcinoma. Journal of clinical pathology *62*, 339-344.

Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868-872.

Lombardo, Y., Scopelliti, A., Cammareri, P., Todaro, M., Iovino, F., Ricci-Vitiani, L., Gulotta, G., Dieli, F., de Maria, R., and Stassi, G. (2011). Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. Gastroenterology *140*, 297-309.

Lou, Y., McDonald, P.C., Oloumi, A., Chia, S., Ostlund, C., Ahmadi, A., Kyle, A., Auf dem Keller, U., Leung, S., Huntsman, D.*, et al.* (2011). Targeting tumor hypoxia: suppression of breast tumor growth and metastasis by novel carbonic anhydrase IX inhibitors. Cancer research *71*, 3364-3376.

Lowry, J.A., and Atchley, W.R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J Mol Evol *50*, 103-115.

Lugli, A., Iezzi, G., Hostettler, I., Muraro, M.G., Mele, V., Tornillo, L., Carafa, V., Spagnoli, G., Terracciano, L., and Zlobec, I. (2010). Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. British journal of cancer *103*, 382-390.

Lugli, A., Karamitopoulou, E., and Zlobec, I. (2012). Tumour budding: a promising parameter in colorectal cancer. British journal of cancer *106*, 1713-1717.

Lugli, A., Spichtin, H., Maurer, R., Mirlacher, M., Kiefer, J., Huusko, P., Azorsa, D., Terracciano, L., Sauter, G., Kallioniemi, O.P.*, et al.* (2005). EphB2 expression across 138 human tumor types in a tissue microarray: high levels of expression in gastrointestinal cancers. ClinCancer Res *11*, 6450-6458.

Lynch, H.T., and de la Chapelle, A. (2003). Hereditary colorectal cancer. The New England journal of medicine *348*, 919-932.

Maeda, M., Ohashi, K., and Ohashi-Kobayashi, A. (2005). Further extension of mammalian GATA-6. Development, growth & differentiation *47*, 591-600.

Maes, C., Carmeliet, G., and Schipani, E. (2012). Hypoxia-driven pathways in bone development, regeneration and disease. Nature reviews Rheumatology *8*, 358-366.

Mahon, P.C., Hirota, K., and Semenza, G.L. (2001). FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. Genes & development *15*, 2675-2686.

Maitra, M., Koenig, S.N., Srivastava, D., and Garg, V. (2010). Identification of GATA6 sequence variants in patients with congenital heart defects. Pediatric research *68*, 281-285.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M.*, et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Mani, S.A., Yang, J., Brooks, M., Schwaninger, G., Zhou, A., Miura, N., Kutok, J.L., Hartwell, K., Richardson, A.L., and Weinberg, R.A. (2007). Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. Proceedings of the National Academy of Sciences of the United States of America *104*, 10069-10074.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7 Suppl 1*, S7.

Markowitz, S.D., and Bertagnolli, M.M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. The New England journal of medicine *361*, 2449-2460.

Marshman, E., Booth, C., and Potten, C.S. (2002). The intestinal epithelial stem cell. BioEssays : news and reviews in molecular, cellular and developmental biology *24*, 91-98.

Massague, J., and Wotton, D. (2000). Transcriptional control by the TGF-beta/Smad signaling system. The EMBO journal *19*, 1745-1754.

Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature *399*, 271-275.

McCormack, N., Molloy, E.L., and O'Dea, S. (2013). Bone morphogenetic proteins enhance an epithelial-mesenchymal transition in normal airway epithelial cells during restitution of a disrupted epithelium. Respiratory research *14*, 36.

McDonald, P.C., Winum, J.Y., Supuran, C.T., and Dedhar, S. (2012). Recent developments in targeting carbonic anhydrase IX for cancer therapeutics. Oncotarget *3*, 84-97.

McIntyre, A., Patiar, S., Wigfield, S., Li, J.L., Ledaki, I., Turley, H., Leek, R., Snell, C., Gatter, K., Sly, W.S.*, et al.* (2012). Carbonic anhydrase IX promotes tumor growth and necrosis in vivo and

inhibition enhances anti-VEGF therapy. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 3100-3111.

Merika, M., and Orkin, S.H. (1993). DNA-binding specificity of GATA family transcription factors. Molecular and cellular biology *13*, 3999-4010.

Merlos-Suarez, A., Barriga, F.M., Jung, P., Iglesias, M., Cespedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Munoz, P*., et al.* (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell stem cell *8*, 511-524.

Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. Proceedings of the National Academy of Sciences of the United States of America *92*, 6991-6995.

Minchenko, A., Leshchinsky, I., Opentanova, I., Sang, N., Srinivas, V., Armstead, V., and Caro, J. (2002). Hypoxia-inducible factor-1-mediated expression of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3 (PFKFB3) gene. Its possible role in the Warburg effect. The Journal of biological chemistry *277*, 6183-6187.

Mir, R., Pradhan, S.J., and Galande, S. (2012). Chromatin organizer SATB1 as a novel molecular target for cancer therapy. Current drug targets *13*, 1603-1615.

Mokry, M., Hatzis, P., de Bruijn, E., Koster, J., Versteeg, R., Schuijers, J., van de Wetering, M., Guryev, V., Clevers, H., and Cuppen, E. (2010). Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. PloS one *5*, e15092.

Mokry, M., Hatzis, P., Schuijers, J., Lansu, N., Ruzius, F.P., Clevers, H., and Cuppen, E. (2012). Integrated genome-wide analysis of transcription factor occupancy, RNA polymerase II binding and steady-state RNA levels identify differentially regulated functional gene classes. Nucleic Acids Res *40*, 148-158.

Molkentin, J.D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. The Journal of biological chemistry *275*, 38949-38952.

Molkentin, J.D., Lin, Q., Duncan, S.A., and Olson, E.N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes & development *11*, 1061-1072.

Molkentin, J.D., Tymitz, K.M., Richardson, J.A., and Olson, E.N. (2000). Abnormalities of the genitourinary tract in female mice lacking GATA5. Molecular and cellular biology *20*, 5256-5260.

Montgomery, R.K., Carlone, D.L., Richmond, C.A., Farilla, L., Kranendonk, M.E., Henderson, D.E., Baffour-Awuah, N.Y., Ambruzs, D.M., Fogli, L.K., Algra, S*., et al.* (2011). Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proceedings of the National Academy of Sciences of the United States of America *108*, 179-184.

Moon, R.T., Bowerman, B., Boutros, M., and Perrimon, N. (2002). The promise and perils of Wnt signaling through beta-catenin. Science *296*, 1644-1646.

Moon, R.T., Kohn, A.D., De Ferrari, G.V., and Kaykas, A. (2004). WNT and beta-catenin signalling: diseases and therapies. Nature reviews Genetics *5*, 691-701.

Morel, A.P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. PloS one *3*, e2888.

Moreno-Bueno, G., Portillo, F., and Cano, A. (2008). Transcriptional regulation of cell polarity in EMT and cancer. Oncogene *27*, 6958-6969.

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.

Morrisey, E.E., Ip, H.S., Lu, M.M., and Parmacek, M.S. (1996). GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. Developmental biology *177*, 309-322.

Morrisey, E.E., Ip, H.S., Tang, Z., Lu, M.M., and Parmacek, M.S. (1997). GATA-5: a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. Developmental biology *183*, 21-36.

Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S., and Parmacek, M.S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & development *12*, 3579-3590.

Mueller, T.D., and Nickel, J. (2012). Promiscuity and specificity in BMP receptor activation. FEBS letters *586*, 1846-1859.

Müller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004). Optimal sample size for multiple testing: the case of gene expression microarrays. Journal of the American Statistical Association *99*, 990-1001.

Munoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S*., et al.* (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. The EMBO journal *31*, 3079-3091.

Murakami, G., Watabe, T., Takaoka, K., Miyazono, K., and Imamura, T. (2003). Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads. Molecular biology of the cell *14*, 2809-2817.

Myat, M.M., and Andrew, D.J. (2000). Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. Development *127*, 4217-4226.

Nakamura, T., Mitomi, H., Kikuchi, S., Ohtani, Y., and Sato, K. (2005). Evaluation of the usefulness of tumor budding on the prediction of metastasis to the lung and liver after curative excision of colorectal cancer. Hepato-gastroenterology *52*, 1432-1435.

Nakanishi, Y., Seno, H., Fukuoka, A., Ueo, T., Yamaga, Y., Maruno, T., Nakanishi, N., Kanda, K., Komekado, H., Kawada, M.*, et al.* (2013). Dclk1 distinguishes between tumor and normal stem cells in the intestine. Nature genetics *45*, 98-103.

Negi, L.M., Talegaonkar, S., Jaggi, M., Ahmad, F.J., Iqbal, Z., and Khar, R.K. (2012). Role of CD44 in tumour progression and strategies for targeting. Journal of drug targeting *20*, 561-573.

Neumann, J., Horst, D., Kriegl, L., Maatz, S., Engel, J., Jung, A., and Kirchner, T. (2012). A simple immunohistochemical algorithm predicts the risk of distant metastases in right-sided colon cancer. Histopathology *60*, 416-426.

Neutra, M.R. (1998). Current concepts in mucosal immunity. V Role of M cells in transepithelial transport of antigens and pathogens to the mucosal immune system. The American journal of physiology *274*, G785-791.

Newcomb, P.A., Norfleet, R.G., Storer, B.E., Surawicz, T.S., and Marcus, P.M. (1992). Screening sigmoidoscopy and colorectal cancer mortality. Journal of the National Cancer Institute *84*, 1572-1575.

Nguyen, L.V., Vanner, R., Dirks, P., and Eaves, C.J. (2012). Cancer stem cells: an evolving concept. Nature reviews Cancer *12*, 133-143.

Nieto, M.A. (2002). The snail superfamily of zinc-finger transcription factors. Nature reviews Molecular cell biology *3*, 155-166.

Notani, D., Gottimukkala, K.P., Jayani, R.S., Limaye, A.S., Damle, M.V., Mehta, S., Purbey, P.K., Joseph, J., and Galande, S. (2010). Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. PLoS biology *8*, e1000296.

Nusse, R., and Varmus, H.E. (1992). Wnt genes. Cell *69*, 1073-1087.

O'Brien, C.A., Kreso, A., Ryan, P., Hermans, K.G., Gibson, L., Wang, Y., Tsatsanis, A., Gallinger, S., and Dick, J.E. (2012). ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. Cancer cell *21*, 777-792.

O'Brien, C.A., Pollett, A., Gallinger, S., and Dick, J.E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature *445*, 106-110.

Ohkubo, T., and Ozawa, M. (2004). The transcription factor Snail downregulates the tight junction components independently of E-cadherin downregulation. Journal of cell science *117*, 1675-1685.

Okada, H., Danoff, T.M., Kalluri, R., and Neilson, E.G. (1997). Early role of Fsp1 in epithelial-mesenchymal transformation. The American journal of physiology *273*, F563-574.

Onichtchouk, D., Chen, Y.G., Dosch, R., Gawantka, V., Delius, H., Massague, J., and Niehrs, C. (1999). Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. Nature *401*, 480-485.

Orkin, S.H. (1998). Embryonic stem cells and transgenic mice in the study of hematopoiesis. The International journal of developmental biology *42*, 927-934.

Oshima, M., Dinchuk, J.E., Kargman, S.L., Oshima, H., Hancock, B., Kwong, E., Trzaskos, J.M., Evans, J.F., and Taketo, M.M. (1996). Suppression of intestinal polyposis in Apc delta716 knockout mice by inhibition of cyclooxygenase 2 (COX-2). Cell *87*, 803-809.

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K.*, et al.* (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427-431.

Pearce, J.J., Penny, G., and Rossant, J. (1999). A mouse cerberus/Dan-related gene family. Developmental biology *209*, 98-110.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nature reviews Cancer *7*, 415-428.

Pennacchietti, S., Michieli, P., Galluzzo, M., Mazzone, M., Giordano, S., and Comoglio, P.M. (2003). Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. Cancer cell *3*, 347-361.

Peterkin, T., Gibson, A., and Patient, R. (2007). Redundancy and evolution of GATA factor requirements in development of the myocardium. Developmental biology *311*, 623-635.

Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J., and Skarnes, W.C. (2000). An LDL-receptor-related protein mediates Wnt signalling in mice. Nature *407*, 535-538.

Pinto, D., Gregorieff, A., Begthel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes & development *17*, 1709-1713.

Polyak, K., and Weinberg, R.A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nature reviews Cancer *9*, 265-273.

Potten, C.S. (1977). Extreme sensitivity of some intestinal crypt cells to X and gamma irradiation. Nature *269*, 518-521.

Pouyssegur, J., Dayan, F., and Mazure, N.M. (2006). Hypoxia signalling in cancer and approaches to enforce tumour regression. Nature *441*, 437-443.

Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E.*, et al.* (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. Cell *149*, 146-158.

Prall, F. (2007). Tumour budding in colorectal carcinoma. Histopathology *50*, 151-162.

Prall, F., Nizze, H., and Barten, M. (2005). Tumour budding as prognostic factor in stage I/II colorectal carcinoma. Histopathology *47*, 17-24.

Pries, A.R., Cornelissen, A.J., Sloot, A.A., Hinkeldey, M., Dreher, M.R., Hopfner, M., Dewhirst, M.W., and Secomb, T.W. (2009). Structural adaptation and heterogeneity of normal and tumor microvascular networks. PLoS computational biology *5*, e1000394.

Prince, M.E., Sivanandan, R., Kaczorowski, A., Wolf, G.T., Kaplan, M.J., Dalerba, P., Weissman, I.L., Clarke, M.F., and Ailles, L.E. (2007). Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proceedings of the National Academy of Sciences of the United States of America *104*, 973-978.

Pyke, C., Ralfkiaer, E., Ronne, E., Hoyer-Hansen, G., Kirkeby, L., and Dano, K. (1994). Immunohistochemical detection of the receptor for urokinase plasminogen activator in human colon cancer. Histopathology *24*, 131-138.

Pyke, C., Salo, S., Ralfkiaer, E., Romer, J., Dano, K., and Tryggvason, K. (1995). Laminin-5 is a marker of invading cancer cells in some human carcinomas and is coexpressed with the receptor for urokinase plasminogen activator in budding cancer cells in colon adenocarcinomas. Cancer research *55*, 4132-4139.

Radtke, F., and Clevers, H. (2005). Self-renewal and cancer of the gut: two sides of a coin. Science *307*, 1904-1909.

Rapisarda, A., and Melillo, G. (2012). Overcoming disappointing results with antiangiogenic therapy by targeting hypoxia. Nature reviews Clinical oncology *9*, 378-390.

Rastaldi, M.P., Ferrario, F., Giardino, L., Dell'Antonio, G., Grillo, C., Grillo, P., Strutz, F., Muller, G.A., Colasanti, G., and D'Amico, G. (2002). Epithelial-mesenchymal transition of tubular epithelial cells in human renal biopsies. Kidney international *62*, 137-146.

Reuter, R. (1994). The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. Development *120*, 1123-1135.

Reuter, R., Grunewald, B., and Leptin, M. (1993). A role for the mesoderm in endodermal migration and morphogenesis in Drosophila. Development *119*, 1135-1145.

Reuter, R., and Leptin, M. (1994). Interacting functions of snail, twist and huckebein during the early development of germ layers in Drosophila. Development *120*, 1137-1150.

Revenu, C., and Gilmour, D. (2009). EMT 2.0: shaping epithelia through collective migration. Current opinion in genetics & development *19*, 338-342.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-111.

Ricci-Vitiani, L., Lombardi, D.G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C., and De Maria, R. (2007). Identification and expansion of human colon-cancer-initiating cells. Nature *445*, 111-115.

Rivera-Feliciano, J., Lee, K.H., Kong, S.W., Rajagopal, S., Ma, Q., Springer, Z., Izumo, S., Tabin, C.J., and Pu, W.T. (2006). Development of heart valves requires Gata4 expression in endothelial-derived cells. Development *133*, 3607-3618.

Roh, M.H., and Margolis, B. (2003). Composition and function of PDZ protein complexes during cell polarization. Am J Physiol Renal Physiol *285*, F377-387.

Rong, L., Liu, J., Qi, Y., Graham, A.M., Parmacek, M.S., and Li, S. (2012). GATA-6 promotes cell survival by up-regulating BMP-2 expression during embryonic stem cell differentiation. Molecular biology of the cell *23*, 3754-3763.

Rossell, D., Guerra, R., and Scott, C. (2008). Semi-parametric differential expression analysis via partial mixture estimation. Statistical applications in genetics and molecular biology *7*, Article15.

Rothenberg, M.E., Nusse, Y., Kalisky, T., Lee, J.J., Dalerba, P., Scheeren, F., Lobo, N., Kulkarni, S., Sim, S., Qian, D.*, et al.* (2012). Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice. Gastroenterology *142*, 1195-1205 e1196.

Ruan, K., Song, G., and Ouyang, G. (2009). Role of hypoxia in the hallmarks of human cancer. Journal of cellular biochemistry *107*, 1053-1062.

Saarnio, J., Parkkila, S., Parkkila, A.K., Haukipuro, K., Pastorekova, S., Pastorek, J., Kairaluoma, M.I., and Karttunen, T.J. (1998). Immunohistochemical study of colorectal tumors for expression of a novel transmembrane carbonic anhydrase, MN/CA IX, with potential value as a marker of cell proliferation. The American journal of pathology *153*, 279-285.

Sakai, Y., Nakagawa, R., Sato, R., and Maeda, M. (1998). Selection of DNA binding sites for human transcriptional regulator GATA-6. Biochemical and biophysical research communications *250*, 682-688.

Salceda, S., and Caro, J. (1997). Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. The Journal of biological chemistry *272*, 22642-22647.

Samad, T.A., Rebbapragada, A., Bell, E., Zhang, Y., Sidis, Y., Jeong, S.J., Campagna, J.A., Perusini, S., Fabrizio, D.A., Schneyer, A.L.*, et al.* (2005). DRAGON, a bone morphogenetic protein co-receptor. The Journal of biological chemistry *280*, 14122-14129.

Sancho, E., Batlle, E., and Clevers, H. (2004). Signaling pathways in intestinal development and cancer. Annual review of cell and developmental biology *20*, 695-723.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes & development *21*, 436-449.

Sandmann, T., Jakobsen, J.S., and Furlong, E.E. (2006). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nature protocols *1*, 2839-2855.

Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nature genetics *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 & development *18*, 1385-1390.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2010). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415-418.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J.*, et al.* (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262-265.

Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. Methods *14*, 381-392.

Schatzkin, A., Freedman, L.S., Dawsey, S.M., and Lanza, E. (1994). Interpreting precursor studies: what polyp trials tell us about large-bowel cancer. Journal of the National Cancer Institute *86*, 1053-1057.

Schepers, A.G., Snippert, H.J., Stange, D.E., van den Born, M., van Es, J.H., van de Wetering, M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science *337*, 730-735.

Scherner, O., Meurer, S.K., Tihaa, L., Gressner, A.M., and Weiskirchen, R. (2007). Endoglin differentially modulates antagonistic transforming growth factor-beta1 and BMP-7 signaling. The Journal of biological chemistry *282*, 13934-13943.

Scholer-Dahirel, A., Schlabach, M.R., Loo, A., Bagdasarian, L., Meyer, R., Guo, R., Woolfenden, S., Yu, K.K., Markovits, J., Killary, K.*, et al.* (2011). Maintenance of adenomatous polyposis coli (APC)-mutant colorectal cancer is dependent on Wnt/beta-catenin signaling. Proceedings of the National Academy of Sciences of the United States of America *108*, 17135-17140.

Schonhoff, S.E., Giel-Moloney, M., and Leiter, A.B. (2004). Minireview: Development and differentiation of gut endocrine cells. Endocrinology *145*, 2639-2644.

Semenza, G.L. (1998). Hypoxia-inducible factor 1: master regulator of O2 homeostasis. Current opinion in genetics & development *8*, 588-594.

Semenza, G.L. (2003). Targeting HIF-1 for cancer therapy. Nature reviews Cancer *3*, 721-732.

Semenza, G.L. (2007). Evaluation of HIF-1 inhibitors as anticancer agents. Drug discovery today *12*, 853-859.

Shaye, D.D., Casanova, J., and Llimargas, M. (2008). Modulation of intracellular trafficking regulates cell intercalation in the Drosophila trachea. Nature cell biology *10*, 964-970.

Sheehan, K.M., Gulmann, C., Eichler, G.S., Weinstein, J.N., Barrett, H.L., Kay, E.W., Conroy, R.M., Liotta, L.A., and Petricoin, E.F., 3rd (2008). Signal pathway profiling of epithelial and stromal compartments of colonic carcinoma reveals epithelial-mesenchymal transition. Oncogene *27*, 323-331.

Shen, F., Li, J., Cai, W., Zhu, G., Gu, W., Jia, L., and Xu, B. (2013). GATA6 predicts prognosis and hepatic metastasis of colorectal cancer. Oncology reports.

Shmelkov, S.V., Butler, J.M., Hooper, A.T., Hormigo, A., Kushner, J., Milde, T., St Clair, R., Baljevic, M., White, I., Jin, D.K*., et al.* (2008). CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. The Journal of clinical investigation *118*, 2111-2120.

Shook, D., and Keller, R. (2003). Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. Mech Dev *120*, 1351-1383.

Shureiqi, I., Zuo, X., Broaddus, R., Wu, Y., Guan, B., Morris, J.S., and Lippman, S.M. (2007). The transcription factor GATA-6 is overexpressed in vivo and contributes to silencing 15-LOX-1 in vitro in human colon cancer. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *21*, 743-753.

Skaer, H. (1989). Cell division in Malpighian tubule development in D. melanogaster is regulated by a single tip cell. Nature *342*, 566.

Skaer, H. (1993). The Alimentary Canal. In Developmental Biology of Drosophila melanogaster, M. Bate, and A. Martinez-Ariaz, eds. (Plainview, NY: Cold Spring Harbor Lab), pp. 941-1012.

Slattery, C., McMorrow, T., and Ryan, M.P. (2006). Overexpression of E2A proteins induces epithelial-mesenchymal transition in human renal proximal tubular epithelial cells suggesting a potential role in renal fibrosis. FEBS letters *580*, 4021-4030.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Snippert, H.J., Schepers, A.G., Delconte, G., Siersema, P.D., and Clevers, H. (2011). Slide preparation for single-cell-resolution imaging of fluorescent proteins in their three-dimensional near-native environment. Nature protocols *6*, 1221-1228.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D*., et al.* (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134-144.

Snippert, H.J., van Es, J.H., van den Born, M., Begthel, H., Stange, D.E., Barker, N., and Clevers, H. (2009). Prominin-1/CD133 marks stem cells and early progenitors in mouse small intestine. Gastroenterology *136*, 2187-2194 e2181.

Sobrado, V.R., Moreno-Bueno, G., Cubillo, E., Holt, L.J., Nieto, M.A., Portillo, F., and Cano, A. (2009). The class I bHLH factors E2-2A and E2-2B regulate EMT. Journal of cell science *122*, 1014-1024.

Sodhi, C.P., Li, J., and Duncan, S.A. (2006). Generation of mice harbouring a conditional loss-of-function allele of Gata6. BMC developmental biology *6*, 19.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nature genetics *21*, 70-71.

Sotillos, S., Diaz-Meco, M.T., Caminero, E., Moscat, J., and Campuzano, S. (2004). DaPKC-dependent phosphorylation of Crumbs is required for epithelial cell polarity in Drosophila. J Cell Biol *166*, 549-557.

Souchelnytskyi, S., Nakayama, T., Nakao, A., Moren, A., Heldin, C.H., Christian, J.L., and ten Dijke, P. (1998). Physical and functional interaction of murine and Xenopus Smad7 with bone morphogenetic protein receptors and transforming growth factor-beta receptors. The Journal of biological chemistry *273*, 25364-25370.

Spaderna, S., Schmalhofer, O., Hlubek, F., Berx, G., Eger, A., Merkel, S., Jung, A., Kirchner, T., and Brabletz, T. (2006). A transient, EMT-linked loss of basement membranes indicates metastasis and poor survival in colorectal cancer. Gastroenterology *131*, 830-840.

Srivastava, A., Pastor-Pareja, J.C., Igaki, T., Pagliarini, R., and Xu, T. (2007). Basement membrane remodeling is essential for Drosophila disc eversion and tumor invasion. Proceedings of the National Academy of Sciences of the United States of America *104*, 2721-2726.

Staller, P., Sulitkova, J., Lisztwan, J., Moch, H., Oakeley, E.J., and Krek, W. (2003). Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. Nature *425*, 307-311.

Tabaries, S., Dong, Z., Annis, M.G., Omeroglu, A., Pepin, F., Ouellet, V., Russo, C., Hassanain, M., Metrakos, P., Diaz, Z*., et al.* (2011). Claudin-2 is selectively enriched in and promotes the formation of breast cancer liver metastases through engagement of integrin complexes. Oncogene *30*, 1318-1328.

Tabaries, S., Dupuy, F., Dong, Z., Monast, A., Annis, M.G., Spicer, J., Ferri, L.E., Omeroglu, A., Basik, M., Amir, E*., et al.* (2012). Claudin-2 promotes breast cancer liver metastasis by facilitating tumor cell interactions with hepatocytes. Molecular and cellular biology *32*, 2979-2991.

Tachezy, M., Zander, H., Gebauer, F., Marx, A., Kaifi, J.T., Izbicki, J.R., and Bockhorn, M. (2012). Activated leukocyte cell adhesion molecule (CD166)--its prognostic power for colorectal cancer patients. The Journal of surgical research *177*, e15-20.

Tafreshi, N.K., Bui, M.M., Bishop, K., Lloyd, M.C., Enkemann, S.A., Lopez, A.S., Abrahams, D., Carter, B.W., Vagner, J., Grobmyer, S.R*., et al.* (2012). Noninvasive detection of breast cancer lymph node metastasis using carbonic anhydrases IX and XII targeted imaging probes. Clinical

cancer research : an official journal of the American Association for Cancer Research *18*, 207-219.

Takahashi-Yanaga, F., and Kahn, M. (2010). Targeting Wnt signaling: can we safely eradicate cancer stem cells? Clinical cancer research : an official journal of the American Association for Cancer Research *16*, 3153-3162.

Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M.F., and Taketo, M.M. (1998). Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. Cell *92*, 645-656.

Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science *334*, 1420-1424.

Takenaga, K. (2011). Angiogenic signaling aberrantly induced by tumor hypoxia. Frontiers in bioscience *16*, 31-48.

Tamai, K., Semenov, M., Kato, Y., Spokony, R., Liu, C., Katsuyama, Y., Hess, F., Saint-Jeannet, J.P., and He, X. (2000). LDL-receptor-related proteins in Wnt signal transduction. Nature *407*, 530-535.

Tanaka, M., Hashiguchi, Y., Ueno, H., Hase, K., and Mochizuki, H. (2003). Tumor budding at the invasive margin can predict patients at high risk of recurrence after curative surgery for stage II, T3 colon cancer. Diseases of the colon and rectum *46*, 1054-1059.

Tanaka, T. (2012). Development of an inflammation-associated colorectal cancer model and its application for research on carcinogenesis and chemoprevention. Int J Inflam *2012*, 658786.

Tepass, U., Fessler, L.I., Aziz, A., and Hartenstein, V. (1994). Embryonic origin of hemocytes and their relationship to cell death in Drosophila. Development *120*, 1829-1837.

Tepass, U., Gruszynski-DeFeo, E., Haag, T.A., Omatyar, L., Torok, T., and Hartenstein, V. (1996). shotgun encodes Drosophila E-cadherin and is preferentially required during cell rearrangement in the neurectoderm and other morphogenetically active epithelia. Genes & development *10*, 672-685.

Tepass, U., and Hartenstein, V. (1994a). The development of cellular junctions in the Drosophila embryo. Developmental biology *161*, 563-596.

Tepass, U., and Hartenstein, V. (1994b). Epithelium formation in the Drosophila midgut depends on the interaction of endoderm and mesoderm. Development *120*, 579-590.

Tepass, U., Theres, C., and Knust, E. (1990). crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila epithelial cells and required for organization of epithelia. Cell *61*, 787-799.

Thiagalingam, S., Lengauer, C., Leach, F.S., Schutte, M., Hahn, S.A., Overhauser, J., Willson, J.K., Markowitz, S., Hamilton, S.R., Kern, S.E*., et al.* (1996). Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. Nature genetics *13*, 343-346.

Thiery, J.P. (2002). Epithelial-mesenchymal transitions in tumour progression. Nature reviews Cancer *2*, 442-454.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thomas, J.T., Canelos, P., Luyten, F.P., and Moos, M., Jr. (2009). Xenopus SMOC-1 Inhibits bone morphogenetic protein signaling downstream of receptor binding and is essential for postgastrulation development in Xenopus. The Journal of biological chemistry *284*, 18994-19005.

Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature *478*, 255-259.

Tomlinson, I.P., Carvajal-Carmona, L.G., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Palles, C., Broderick, P., Jaeger, E.E., Farrington, S*., et al.* (2011a). Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS genetics *7*, e1002105.

Tomlinson, I.P., Carvajal-Carmona, L.G., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Palles, C., Broderick, P., Jaeger, E.E., Farrington, S*., et al.* (2011b). Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS genetics *7*, e1002105.

Topol, L.Z., Bardot, B., Zhang, Q., Resau, J., Huillard, E., Marx, M., Calothy, G., and Blair, D.G. (2000). Biosynthesis, post-translation modification, and functional characterization of Drm/Gremlin. The Journal of biological chemistry *275*, 8785-8793.

Tsai, J.H., Donaher, J.L., Murphy, D.A., Chau, S., and Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer cell *22*, 725-736.

Uchida, N., Buck, D.W., He, D., Reitsma, M.J., Masek, M., Phan, T.V., Tsukamoto, A.S., Gage, F.H., and Weissman, I.L. (2000). Direct isolation of human central nervous system stem cells. Proceedings of the National Academy of Sciences of the United States of America *97*, 14720-14725.

Uemura, T., Oda, H., Kraut, R., Hayashi, S., Kotaoka, Y., and Takeichi, M. (1996). Zygotic Drosophila E-cadherin expression is required for processes of dynamic epithelial cell rearrangement in the Drosophila embryo. Genes & development *10*, 659-671.

Ueno, H., Mochizuki, H., Shinto, E., Hashiguchi, Y., Hase, K., and Talbot, I.C. (2002). Histologic indices in biopsy specimens for estimating the probability of extended local spread in patients with rectal carcinoma. Cancer *94*, 2882-2891.

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.* (2002a). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.

van de Wetering, M., Sancho, E., Verweij, C., de, L.W., Oving, I., Hurlstone, A., van der, H.K., Batlle, E., Coudreuse, D., Haramis, A.P.*, et al.* (2002b). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.

van der Flier, L.G., and Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annual review of physiology *71*, 241-260.

van der Flier, L.G., Haegebarth, A., Stange, D.E., van de Wetering, M., and Clevers, H. (2009a). OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. Gastroenterology *137*, 15-17.

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.*, et al.* (2007). The Intestinal Wnt/TCF Signature. Gastroenterology *132*, 628-632.

van der Flier, L.G., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Begthel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009b). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

van Dop, W.A., Uhmann, A., Wijgerde, M., Sleddens-Linkels, E., Heijmans, J., Offerhaus, G.J., van den Bergh Weerman, M.A., Boeckxstaens, G.E., Hommes, D.W., Hardwick, J.C.*, et al.* (2009). Depletion of the colonic epithelial precursor cell compartment upon conditional activation of the hedgehog pathway. Gastroenterology *136*, 2195-2203 e2191-2197.

van Es, J.H., Haegebarth, A., Kujala, P., Itzkovitz, S., Koo, B.K., Boj, S.F., Korving, J., van den Born, M., van Oudenaarden, A., Robine, S.*, et al.* (2012a). A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. Molecular and cellular biology *32*, 1918-1927.

van Es, J.H., Sato, T., van de Wetering, M., Lyubimova, A., Nee, A.N., Gregorieff, A., Sasaki, N., Zeinstra, L., van den Born, M., Korving, J.*, et al.* (2012b). Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. Nature cell biology *14*, 1099-1104.

van Kempen, L.C., van den Oord, J.J., van Muijen, G.N., Weidle, U.H., Bloemers, H.P., and Swart, G.W. (2000). Activated leukocyte cell adhesion molecule/CD166, a marker of tumor progression in primary malignant melanoma of the skin. The American journal of pathology *156*, 769-774.

Varnat, F., Siegl-Cachedenier, I., Malerba, M., Gervaz, P., and Ruiz i Altaba, A. (2010). Loss of WNT-TCF addiction and enhancement of HH-GLI1 signalling define the metastatic transition of human colon carcinomas. EMBO Mol Med *2*, 440-457.

Vasen, H.F., Wijnen, J.T., Menko, F.H., Kleibeuker, J.H., Taal, B.G., Griffioen, G., Nagengast, F.M., Meijers-Heijboer, E.H., Bertario, L., Varesco, L.*, et al.* (1996). Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology *110*, 1020-1027.

Vaupel, P., and Mayer, A. (2007). Hypoxia in cancer: significance and impact on clinical outcome. Cancer metastasis reviews *26*, 225-239.

Vermeulen, L., De Sousa, E.M.F., van der Heijden, M., Cameron, K., de Jong, J.H., Borovski, T., Tuynman, J.B., Todaro, M., Merz, C., Rodermond, H*., et al.* (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nature cell biology *12*, 468-476.

Vermeulen, L., Sprick, M.R., Kemper, K., Stassi, G., and Medema, J.P. (2008). Cancer stem cells--old concepts, new insights. Cell death and differentiation *15*, 947-958.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Developmental cell *19*, 713-726.

Vincan, E., and Barker, N. (2008). The upstream components of the Wnt signalling pathway in the dynamic EMT and MET associated with colorectal cancer progression. Clinical & experimental metastasis *25*, 657-663.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F*., et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Waite, K.A., and Eng, C. (2003). From developmental disorder to heritable cancer: it's all in the BMP/TGF-beta family. Nature reviews Genetics *4*, 763-773.

Wang, G.L., Jiang, B.H., Rue, E.A., and Semenza, G.L. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proceedings of the National Academy of Sciences of the United States of America *92*, 5510-5514.

Wang, X., Zheng, M., Liu, G., Xia, W., McKeown-Longo, P.J., Hung, M.C., and Zhao, J. (2007). Kruppel-like factor 8 induces epithelial to mesenchymal transition and epithelial cell invasion. Cancer research *67*, 7184-7193.

Waterman, M.L. (2004). Lymphoid enhancer factor/T cell factor expression in colorectal cancer. Cancer metastasis reviews *23*, 41-52.

Wenger, R.H., Rolfs, A., Kvietikova, I., Spielmann, P., Zimmermann, D.R., and Gassmann, M. (1997). The mouse gene for hypoxia-inducible factor-1alpha--genomic organization, expression and characterization of an alternative first exon and 5' flanking sequence. European journal of biochemistry / FEBS *246*, 155-165.

Will, B., Vogler, T.O., Bartholdy, B., Garrett-Bakelman, F., Mayer, J., Barreyro, L., Pandolfi, A., Todorova, T.I., Okoye-Okafor, U.C., Stanley, R.F*., et al.* (2013). Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nature immunology *14*, 437-445.

Wilson, W.R., and Hay, M.P. (2011). Targeting hypoxia in cancer therapy. Nature reviews Cancer *11*, 393-410.

Wodarz, A., and Nusse, R. (1998). Mechanisms of Wnt signaling in development. Annual review of cell and developmental biology *14*, 59-88.

Xu, B., Chen, C., Chen, H., Zheng, S.G., Bringas, P., Jr., Xu, M., Zhou, X., Chen, D., Umans, L., Zwijsen, A.*, et al.* (2011). Smad1 and its target gene Wif1 coordinate BMP and Wnt signaling activities to regulate fetal lung development. Development *138*, 925-935.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927-939.

Yang, J., and Weinberg, R.A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Developmental cell *14*, 818-829.

Yee, D.S., Tang, Y., Li, X., Liu, Z., Guo, Y., Ghaffar, S., McQueen, P., Atreya, D., Xie, J., Simoneau, A.R.*, et al.* (2010). The Wnt inhibitory factor 1 restoration in prostate cancer cells was associated with reduced tumor growth, decreased capacity of cell migration and invasion and a reversal of epithelial to mesenchymal transition. Molecular cancer *9*, 162.

Yin, A.H., Miraglia, S., Zanjani, E.D., Almeida-Porada, G., Ogawa, M., Leary, A.G., Olweus, J., Kearney, J., and Buck, D.W. (1997). AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood *90*, 5002-5012.

Yip, M.L., Lamka, M.L., and Lipshitz, H.D. (1997). Control of germ-band retraction in Drosophila by the zinc-finger protein HINDSIGHT. Development *124*, 2129-2141.

Yu, F., White, S.B., Zhao, Q., and Lee, F.S. (2001). HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. Proceedings of the National Academy of Sciences of the United States of America *98*, 9630-9635.

Yu, P.B., Deng, D.Y., Lai, C.S., Hong, C.C., Cuny, G.D., Bouxsein, M.L., Hong, D.W., McManus, P.M., Katagiri, T., Sachidanandan, C.*, et al.* (2008a). BMP type I receptor inhibition reduces heterotopic [corrected] ossification. Nature medicine *14*, 1363-1369.

Yu, P.B., Hong, C.C., Sachidanandan, C., Babitt, J.L., Deng, D.Y., Hoyng, S.A., Lin, H.Y., Bloch, K.D., and Peterson, R.T. (2008b). Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. Nature chemical biology *4*, 33-41.

Yui, S., Nakamura, T., Sato, T., Nemoto, Y., Mizutani, T., Zheng, X., Ichinose, S., Nagaishi, T., Okamoto, R., Tsuchiya, K.*, et al.* (2012). Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5(+) stem cell. Nature medicine *18*, 618-623.

Zeisberg, M., Shah, A.A., and Kalluri, R. (2005). Bone morphogenic protein-7 induces mesenchymal to epithelial transition in adult renal fibroblasts and facilitates regeneration of injured kidney. The Journal of biological chemistry *280*, 8094-8100.

Zhang, S., Fei, T., Zhang, L., Zhang, R., Chen, F., Ning, Y., Han, Y., Feng, X.H., Meng, A., and Chen, Y.G. (2007). Smad7 antagonizes transforming growth factor beta signaling in the nucleus by interfering with functional Smad-DNA complex formation. Molecular and cellular biology *27*, 4488-4499.

Zhang, Y., Goss, A.M., Cohen, E.D., Kadzik, R., Lepore, J.J., Muthukumaraswamy, K., Yang, J., DeMayo, F.J., Whitsett, J.A., Parmacek, M.S*., et al.* (2008). A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. Nature genetics *40*, 862-870.

Zhong, Y., Wang, Z., Fu, B., Pan, F., Yachida, S., Dhara, M., Albesiano, E., Li, L., Naito, Y., Vilardell, F*., et al.* (2011). GATA6 activates Wnt signaling in pancreatic cancer by negatively regulating the Wnt antagonist Dickkopf-1. PloS one *6*, e22129.

Zhu, H., Kavsak, P., Abdollah, S., Wrana, J.L., and Thomsen, G.H. (1999). A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. Nature *400*, 687-693.

Ziskin, J.L., Dunlap, D., Yaylaoglu, M., Fodor, I.K., Forrest, W.F., Patel, R., Ge, N., Hutchins, G.G., Pine, J.K., Quirke, P*., et al.* (2012). In situ validation of an intestinal stem cell signature in colorectal cancer. Gut.

Zlobec, I., and Lugli, A. (2010). Epithelial mesenchymal transition and tumor budding in aggressive colorectal cancer: tumor budding as oncotarget. Oncotarget *1*, 651-661.

Zohn, I.E., Li, Y., Skolnik, E.Y., Anderson, K.V., Han, J., and Niswander, L. (2006). p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell *125*, 957-969.