ABSTRACT

Title of Dissertation:	DIVERGENT ROLES OF THE TCF4 AND LEF1 WNT SIGNALING TRANSCRIPTION FACTORS IN COLON CANCER
	Markus Alexander Brown, Doctor of Philosophy, 2021
Dissertation directed by:	Professor, Kan Cao, Department of Cell Biology & Molecular Genetics Senior Investigator, Thomas Ried, Genetics Branch, Center for Cancer Research, NCI, NIH

The canonical WNT signaling pathway is necessary for guiding cell growth during embryonic development. In adults, WNT signaling maintains tissue stem cells and therefore plays an essential role in tissue homeostasis. In the colon, the WNT transcription factor, TCF4, is necessary for maintaining the intestinal stem cells. The initiating event in colon cancer is the aberrant activation of the WNT signaling pathway, which results in constitutive activity of TCF4. To determine how TCF4 influences colon cancer cell behavior, we silenced *TCF7L2*, the gene encoding TCF4, and used RNA sequencing and Hi-C to measure changes in transcription and nuclear structure in the SW480 colon cancer cell line. Loss of TCF4 resulted in A/B compartment switching, local chromatin reorganization, and a dramatic up-regulation in transcription.

However, A/B compartment switching was not associated with changes in gene expression. We also found that loss of TCF4 resulted in the up-regulation of *LEF1*, another WNT transcription factor. Expressed LEF1 isoforms were found to be transcriptionally competent and over-compensated for WNT signaling activity upon loss of TCF4, suggesting a WNT-intrinsic feedback mechanism. Over-expression of *LEF1* altered WNT signaling output to favor the expression of lymphoid genes, as opposed to a TCF4-based transcriptional program. ChIP-seq demonstrated that TCF4 and LEF1 bind distinct target genes, though they synergize to express *MYC*. TCF4 was found to bind the *LEF1* promoter, indicative of direct repression, though LEF1 did not bind the *TCF7L2* promoter. The CtBP1 protein, a known binding partner of TCF4, was found to be the most potent repressor of *LEF1* expression. This demonstrates that despite the overall activation of WNT signaling in colon cancer, repressive functions of the WNT transcription factors are still intact, and the repression of *LEF1* by TCF4 maintains a TCF4-centric transcriptional program in colon cancer cells.

DIVERGENT ROLES OF THE TCF4 AND LEF1 WNT SIGNALING TRANSCRIPTION FACTORS IN COLON CANCER

by

Markus Alexander Brown

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Kan Cao, Chair
Adjunct Professor Thomas Ried, Co-Chair
Professor Eric Haag, Dean's Representative
Professor Antony Jose
Professor Sougata Roy

Dedication

To Dr. Theodore A. Napora, for instilling within me an interest in the sciences, as well as for his generosity, which made obtaining my doctorate far more comfortable than it should have been.

To my younger brother, Benny, for demonstrating what courage and resilience really mean.

To my parents, for their continued support, and for only infrequently asking, “How much longer?”

Acknowledgements

I thank Dr. Thomas Ried for his continued support, guidance, tolerance, and generosity, as he supervised me from my infancy as a high school student up until the present. His kindness and trust have significantly influenced the scientist that I have become.

I thank Dr. Kan Cao for her support, kindness, and understanding as my advisor. Had it not been for her insight and guidance, I would still be attempting to earn my PhD.

I thank the members of my committee (past and present), Drs. Sridhar Hannenhalli, Najib El-Sayed, Eytan Ruppin, Eric Haag, Antony Jose, and Sougata Roy for their time, insight, and constructive feedback.

I thank Dr. Michael Boutros of the DKFZ in Heidelberg, Germany, for graciously allowing me to perform experiments in his lab for six months and for the opportunity to join in the lab's ATCG meetings for useful discussions regarding our experimental results. I also thank Dr. Tianzuo Zhan for his guidance and supervision during my stay at the DKFZ.

I thank Dr. Indika Rajapakse, for allowing me to visit his laboratory on three separate occasions, and Dr. Haiming Chen, for teaching me Hi-C. I am also thankful for the help of Gabrielle Dotson and Scott Ronquist as they performed the analysis of the Hi-C data, a feat beyond my capabilities, as well as for their patience in explaining how the analysis was performed.

I thank Dr. Raymond St. Leger for sharing his seemingly endless expanse of knowledge with me, as well as for his humor, generosity, and patience, which made our many discussions all the more enjoyable.

I thank Dr. Yuri Lazebnik for his insight and penetrating questions regarding my project as well as for his general support in all things science related.

I thank the members of the Riedlab for their guidance during the past decade of my life, particularly for their patience and endurance in dealing with a high school student, who frankly, did not know much of anything at all, and still doesn't. I applaud them for determining the sole method of action by which they could finally rid themselves of me (helping with my dissertation). A special thank you to Drs. Darawalee Wangsa, Jordi Camps Polo, and Chanelle Case, who have played a fundamental role in my development into a fledgling scientist. I thank Dr. Wei-Dong Chen for his abundant knowledge regarding cloning techniques, and Dr. Yu Hue for his insight into bioinformatic analysis. I thank Dr. Daniela Hirsch for sharing her extensive knowledge regarding all things

biological with me and for always sharing her information willingly. I thank Dr. Georg Emons for his insight regarding WNT signaling and for the particularly brilliant idea of suggesting we go to a monster truck rally, as his son was into that at the time. I thank Valentin Rausch for taking me under his wing in my youth, for the many visits to Five Guys, and for saving me from suffering the consequences of my mischief. I thank Drs. Danny Wangsa, Kerstin Heselmeyer-Haddad, and Rachel Lokanga, as well as Irianna Torres, for maintaining an enjoyable, relaxed laboratory atmosphere – the most important factor of all – as well as for their general insight and advice. A most sincere thank you to the students I have had the privilege of working with over the years – Janina Lehmann, Hannah Warshawsky, and Juliane Kamlah – to name a few, for their time, their courage in trusting someone who wasn't always sure of what the best way forward was, and for nevertheless continuing to work alongside me, I can never thank you enough. I thank Buddy Chen, Mai Tran, and Melissa Shue, for ensuring that reagents were ordered promptly and that I was paid handsomely for my mischief. And an offering of thanks, as well as a request for forgiveness, to all those individuals involved whose names have eluded my swiss cheese memory.

Table of Contents

List of Tables

List of Figures

Figure 17 | WNT TF expression in the Transverse and Sigmoid Colon from GTEx (n=199, 180). The bars show the average and the error bars denote the standard deviation. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

Figure 18 | WNT TF Percent Expression (n=379, 254).

Figure 19 | *TCF7L2* transcript abundance was assayed with qPCR against the reference gene *YWHAZ* (left). Three biological replicates are plotted for each time point, shown as green dots (n=3). The green line represents the average and the shaded ribbon denotes the standard deviation. Statistical significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$). TCF4 abundance was determined via western blot (right) with TBP as loading control.

Figure 20 | Differential gene expression in *TCF7L2*-silenced SW480 cells. The threshold for $\log_2$ fold change was set to two and the threshold for the FDR adjusted *p*-value was set to 0.05. The genes which exceeded the fold change threshold and were below the *p*-value threshold are colored green for up-regulated and red for down-regulated. The number of genes in each category are shown in the corresponding color. Two biological replicates were used (n=2).

Figure 21 | GSEA for *TCF7L2*-silenced SW480 cells. The NESs for all 50 Hallmark gene sets were calculated and the top ten up-regulated as well as the top 10 down-regulated pathways, based on their adjusted *p*-value at the last time point, were plotted.

Figure 22 | Influence of A/B compartment partitioning on gene expression. Gene expression for Chromosome 13 (left) and Chromosome 20 (middle) is shown in blue, while A/B compartments are shown in green and red. Contact matrices are shown on the bottom in red. A/B compartment switching dynamics are shown on the right. A/B compartment switching was clustered into eight distinct groups, represented by roman numerals on the left, with the number of genes falling into each specific pattern shown on the right.

Figure 23 | Expression of the *CEACAM* loci at 0 hours (bottom) and 72 hours (top) with TCF4 and SP1 DNA binding in red and Hi-C contact matrices in pink with calculated TADs denoted by the solid (0 hours) and dotted (72 hours) black lines.

Figure 24 | Up-regulation of *CEACAM1* upon silencing of *TCF7L2* in a select group of colon cancer lines (n=3).

Figure 25 | Gene expression (left) and chromatin structure (right) dynamics for the top ten up-regulated and top ten down-regulated pathway gene sets. Gene expression is ordered by decreasing NES score. Chromatin structure is ordered based on both the Frobenius norm and the order from gene expression.

Figure 26 | WNT transcription factor expression upon silencing of *TCF7L2*. The dots represent biological replicates (n=2), the line represents the average expression, and the shaded ribbon represents the standard deviation.

List of Abbreviations

ACF: Aberrant Crypt Foci

AFAP: Attenuated Familial Adenomatous Polyposis

FAP: Familial Adenomatous Polyposis

JPS: Juvenile Polyposis Syndrome

LEF1: Lymphoid Enhancer Binding Factor 1

MAP: *MUTYH*-Associated Polyposis

MSI: Microsatellite Instable

MSI-H: Microsatellite Instable – High

MSI-L: Microsatellite Instable – Low

MSS: Microsatellite Stable

NSAID: Non-Steroidal Anti-Inflammatory Drug

PJS: Peutz-Jeghers Syndrome

RIN: RNA Integrity Number

TAD: Topologically Associating Domain

TCF: T-Cell Factor

TF: Transcription Factor

TGFβ: Transforming Growth Factor Beta

WNT: A portmanteau of Wg and Int

Chapter 1: Introduction

The canonical WNT signaling pathway is necessary for metazoan life. In the developing embryo, it promotes growth and spatially orients cells into tissues. In adults, WNT signaling maintains adult stem cells and therefore plays an essential role in tissue homeostasis. However, the potent growth inducing capacity of WNT signaling can be corrupted to drive the aberrant growth of cells in cancer.

In mammals, WNT signaling is crucial for the development of the colon in embryos and maintenance of colonic stem cells in adults. The replication of these stem cells is driven by WNT signaling activity and their progeny differentiate to form the various epithelial cell types of the colon, which perform the colonic functions. The growth of stem cells and their progeny is carefully balanced by the elimination of differentiated cells to ensure that cells are actively renewed, but do not overgrow.

It is generally accepted that the initiating mutation in colon cancer occurs in the WNT signaling pathway. These mutations typically occur in *APC*, a key tumor suppressor in the pathway. The result of the mutation is constitutively activated WNT signaling, which results in continuous growth of colonic cells and decreased differentiation. This imbalance in the growth/death ratio results in the formation of a tumor, which may extrude into the intestinal lumen. As these cells continue to divide, they accrue additional mutations, typically in *KRAS*, *SMAD4*, and *TP53*, which enhances their ability to proliferate and imbues them with the capability to metastasize.

Mutations in *APC* lead to constitutive WNT signaling activity by decreasing the ability of *APC* to facilitate the destruction of β-catenin, a potent transcriptional activator. In normal cells, the nuclear levels of β-catenin are very low, and β-catenin is localized primarily in the cytoplasm, where it is actively destroyed. However, in cells with mutated *APC*, β-catenin accumulates in the cytoplasm, which then migrates to the nucleus, resulting in high levels of nuclear β-catenin. Nuclear β-catenin binds to the TCF/LEF family of transcription factors, which are the canonical WNT signaling transcription factors. These factors are not standard transcription factors, as they are incapable of regulating transcription on their own. Instead, their influence on transcription is determined by their binding partner(s). When bound to β-catenin, these transcription factors activate the transcription of WNT target genes, such as *MYC* and *CCND1*, which results in cell cycle progression and growth. Biallelic mutation of *APC* therefore results in constitutive activity through β-catenin/TCF4 complexes.

Humans have four TCF/LEF factors: *TCF7*, *TCF7L1*, *TCF7L2*, and *LEF1*, which encode the TCF1, TCF3, TCF4, and LEF1 proteins, respectively. TCF4 is crucial for development and maintenance of the colon, while the other factors are expendable. TCF1 and LEF1 are involved in the immune system, while TCF3 regulates skin development. In the human colon, *TCF7L1* and *TCF7L2* are abundant, while *TCF7* and *LEF1* are either not detectable, or detectable to a minute degree. In colon cancer, however, *TCF7* and *LEF1* are more abundant and *TCF7L1* expression is lost. This raises the question, what roles do TCF1 and LEF1 play in colon cancer? TCF4 may hold the key, as silencing of *TCF7L2* (TCF4) induces a potent up-regulation of *LEF1*.

Intrigued by this interaction and realizing it had not been explored in the literature, I investigated the interaction between TCF4 and *LEF1*. I sought to answer the questions which I believed most directly addressed our lack of knowledge:

1. **What are the consequences of silencing *TCF7L2* in colon cancer?**
    a. How does gene expression change?
    b. How does nuclear structure change?
    c. Do changes in nuclear structure and gene expression correlate?

2. **What are the consequences of silencing *LEF1* in colon cancer?**
    a. Are the protein isoforms of LEF1 capable of driving transcription?
    b. What pathways are influenced by LEF1?
    c. Does the activity of LEF1 complement the actions of TCF4 or are the target genes non-overlapping?

3. **How does TCF4 regulate *LEF1*?**
    a. Does TCF4 bind a regulatory element of *LEF1*?
    b. Does LEF1 bind a regulatory element of *TCF7L2*?
    c. If TCF4 does bind *LEF1*, which domain of TCF4 is responsible for repressing *LEF1* (which is then reduced upon silencing of TCF4, resulting in the observed up-regulation)?

Chapter 2: Literature Review

2.1 Colon Cancer: Epidemiology

2.1.1 The Impact of Colorectal Cancer on Modern Society

Colon cancer refers to those tumors which originate in the colon and whose cell of origin is a colonic, typically epithelial, cell. Given the high similarity in pathogenesis observed between colon and rectal cancers, these two cancer types are often grouped together and are jointly referred to as colorectal cancer. Colorectal cancer accounts for approximately 10% of all newly diagnosed cancer cases and cancer-related deaths worldwide, which translates to approximately 1.8 million new cases and 885,000 deaths per year[1]. It is the third most frequently diagnosed cancer in males and the second in females. In terms of cancer-associated mortality, colorectal cancer ranks fourth in males, and third in females[1]. Men are approximately 30% more likely to develop colorectal cancer than women[1-3]. The global distribution of colorectal cancer is biased towards highly-developed countries, such as the United States and Europe, due to risk factors associated with a western lifestyle including: physical inactivity, diets high in red and processed meats, excess body weight, and smoking[4-6] (Figure 1). Less developed countries are plagued instead by cancers caused by infection[7-9]. However, continuing socioeconomic progress in developing countries results in a shift in cancer type prevalence[10,11]. The annual burden of colorectal cancer worldwide is therefore expected to increase to ~2.2 million cases and 1.1 million deaths by 2030, despite stabilized or decreasing case and mortality rates in highly-developed countries[12,13].

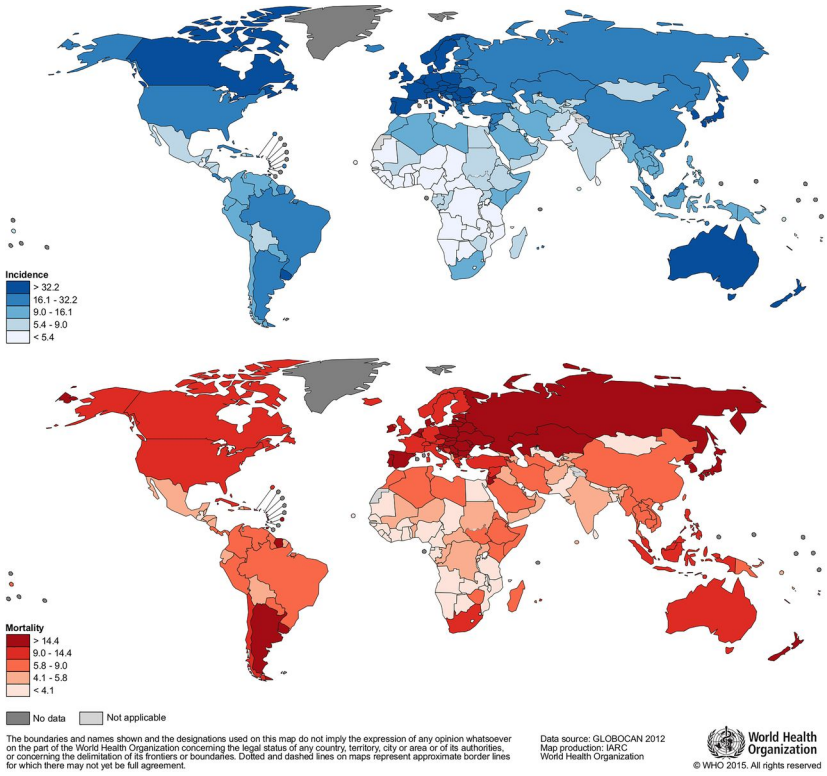

Figure 1 | Worldwide colorectal cancer incidence and mortality rates (age adjusted according to the world standard population, per 100,000) in males (GLOBOCAN 2012). Reproduced with permission from Arnold et al., 2017.

2.1.2 Colorectal Cancer Dynamics in the United States

In the United States, colorectal cancer is the second most frequent cause of cancer death[14]. The likelihood that a man or woman living in the United States will be diagnosed during their lifetime is ~4.2%, and, in 2020, an estimated 147,950 individuals will be diagnosed with colorectal cancer in the United States[14]. The five-year relative survival rate is dependent upon the stage at diagnosis and ranges from 90% for patients diagnosed with early, localized disease to 5% for patients with late, metastatic disease[15,16]. Routine screening is therefore crucial to minimize colorectal cancer associated mortality, by diagnosing the disease at an early, treatable stage[17].

As is commonly observed across cancer types, the rate of colorectal cancer incidence increases dramatically with age. The incidence rate approximately doubles every five years until age 50, at which point there is a 30% increase in the incidence rate for every five-year interval[14]. Counterintuitively, the median age of diagnosis has dropped from 72 years of age in 2001-2002 to 66 years of age in 2015-2016. The reduction is a result of decreasing colorectal cancer rates in those ages 55 and older, coupled with an increase in those younger than 55 years[18-20]. The declining incidence and mortality rates in the elderly population is a result of reduced exposure to risk factors, increased screening, and improved treatment modalities[16]. The factors underlying the 51% increase in colorectal cancer incidence in young adults remain unclear, though likely reflect increased exposure to risk factors, specifically: physical inactivity, diets high in red and processed meats, and obesity[21,22]. As a result of the rising colorectal cancer incidence rates in young adults, the American Cancer Society (ACS) now recommends that patients aged 45, with an average risk of colorectal cancer, undergo regular screening with either a high-sensitivity stool-based test or a colonoscopy[22].

2.1.3 Colorectal Cancer Risk Factors and Inherited Syndromes

Societal and hereditary risk factors both contribute to colorectal cancer incidence. More than one-half of all colorectal cancer cases and deaths have been attributed to societal risk factors, such as those associated with a western lifestyle[21]. Fortunately, these factors are modifiable and include physical inactivity, diets high in processed and red meats, excess body weight, and smoking (Figure 2).

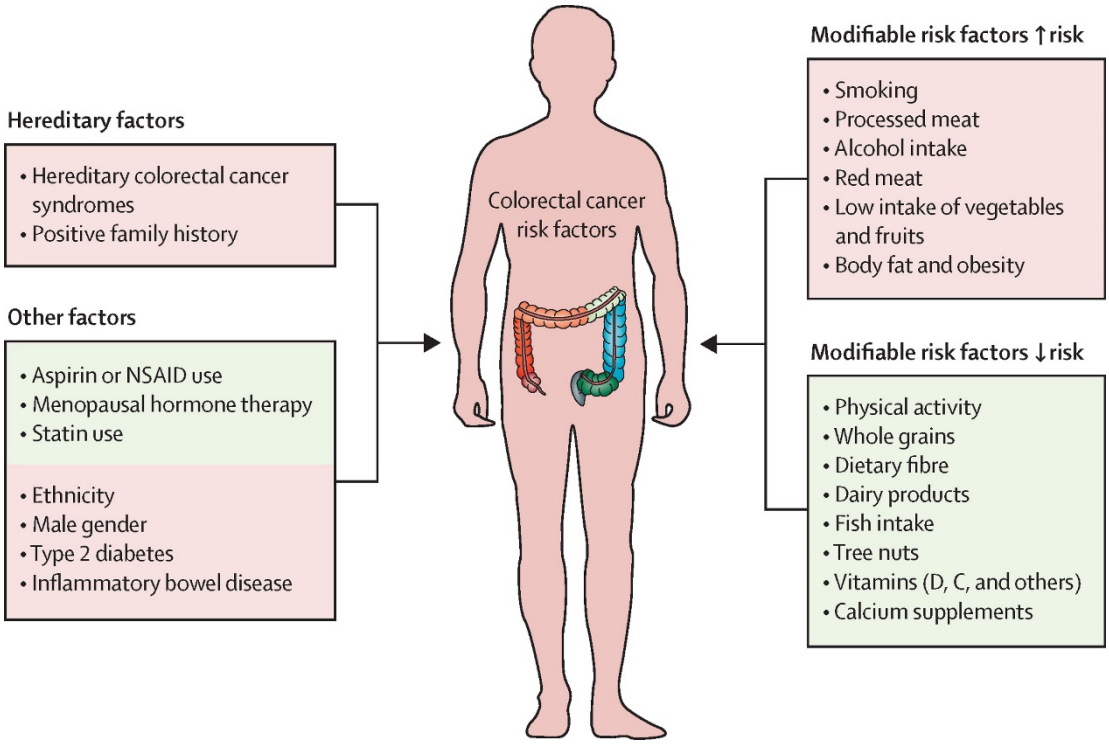

Figure 2 | Colorectal cancer risk factors. Modifiable risk factors which decrease risk are highlighted in green, while modifiable and non-modifiable risk factors which increase risk, are highlighted in red. Reproduced with permission from Dekker et al., 2019.

Hereditary risk factors, which persist in the DNA, account for 10-20% of all colorectal cancer patients, while the heritability of colorectal cancer ranges from 12-35%[23-25]. Approximately 5-10% of colorectal cancer cases are attributed to a well-characterized hereditary colorectal cancer syndrome[26]. Hereditary colorectal cancer syndromes can be divided into polyposis and non-polyposis syndromes. The polyposis syndromes are easily recognized due to the polyp burden in the colon, while non-polyposis syndromes are more difficult to recognize as patients present fewer polyps and these polyps resemble sporadic lesions. The polyposis syndromes can be further subdivided into adenomatous and hamartomatous syndromes. The adenomatous polyposis syndromes include: Familial Adenomatous Polyposis (FAP), Attenuated Familial Adenomatous Polyposis (AFAP), and *MUTYH*-Associated Polyposis (MAP).

The hamartomatous polyposis syndromes include: Peutz-Jeghers Syndrome (PJS), Juvenile Polyposis Syndrome (JPS), Cowden Syndrome, and Serrated/Hyperplastic Polyposis Syndrome. The single non-polyposis syndrome is Lynch Syndrome. Individuals suffering from any of the polyposis or non-polyposis syndromes are recommended to begin earlier screening for the presence of colorectal polyps[26]. The age at which screening should be initiated varies by syndrome, with FAP patients recommended to begin screening during puberty, while Lynch Syndrome patients are recommended to initiate screening between 20-25 years of age[26].

2.2 Colon Cancer: Manifestation

2.2.1 The Human Colon

The colon of *Homo sapiens* is a smooth, elongate, muscular tube approximately four feet in length and three inches in diameter, which links the cecum to the rectum[27]. The colon is divided into four major segments: the ascending colon (which connects to the cecum on the right side of the body), the transverse colon, the descending colon, and the sigmoid colon (which connects to the rectum). The colon can be divided into the right and left colon, with the right colon comprised of the ascending and transverse colon, and the left colon comprised of the descending and sigmoid colon (Figure 3). The purpose of the colon is to extract water from dietary waste, thereby concentrating and forming the stool. Once stool is in the rectum, nerves signal the brain to defecate.

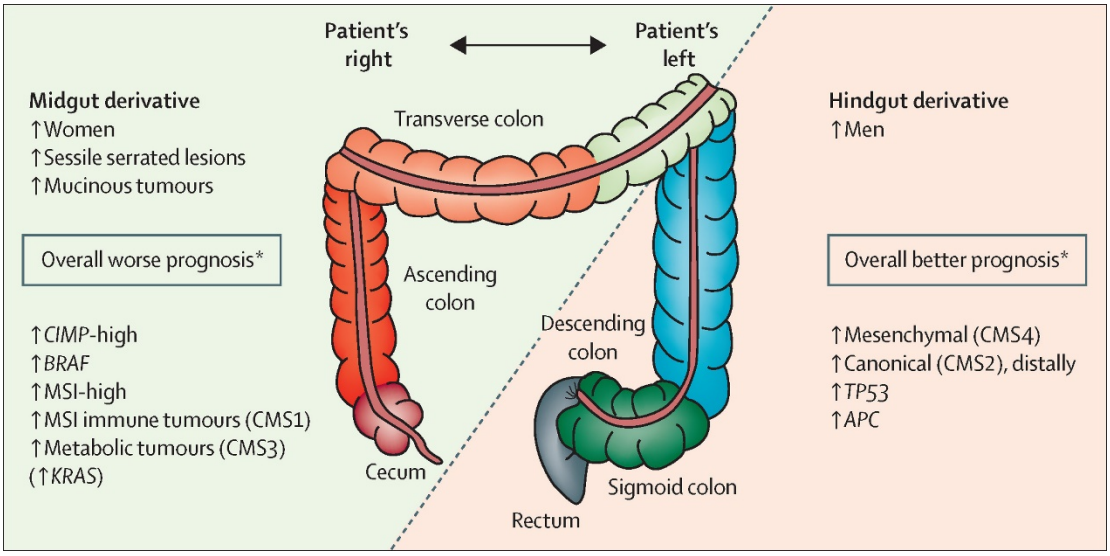

Figure 3 | The regions of the colon. The ascending colon is shown in red, the transverse colon is shown in orange and light green, the descending colon is in blue, while the sigmoid colon is colored green. The colon empties into the rectum, which is shown in gray. Reproduced with permission from Dekker et al., 2019.

Despite fulfilling a single function, the role of the colonic epithelium is paradoxical as it must be permeable to facilitate water absorption, yet impermeable to prevent the vast numbers of microbes which reside within the colon from accessing the body. It must also be structurally resilient to endure the passing stool. The structure of the colon, as well as the function of the colonic epithelium, has been evolutionarily shaped to adjust to these multi-faceted constraints. To facilitate the movement of the stool, the internal surface of the colon is smooth and secretes mucus. Undulating contractions of smooth muscle push the stool forwards. Unlike the small intestine, no villi are found in the colon[28]. If villi were to be found in the colon, likely to facilitate water absorption by increased surface area, they would be damaged by the passing feces, which is the fate of those colonic polyps extruding into the intestinal lumen. Therefore, the strategy of increasing surface area to facilitate absorption utilized by the small intestine is not a viable strategy in the colon. Instead, to facilitate the absorption of water, the colonic epithelium is a single cell layer thick, consisting of cells with channel proteins designed for water transportation. This single layer epithelium is a vulnerability considering the trillions of colon-resident microbes[29]. Unlike the skin, which utilizes multiple layers of dead cells to generate an impermeable barrier, this strategy is not viable in the colon due to its requirement for absorption. Additionally, given the constant movement of the stool, small abrasions which damage the epithelial cell layer could create openings for opportunistic infection. To account for the weak structural integrity of a single cell layer, these epithelial cells are constantly being renewed. Indeed, the intestinal stem cells, which are the progenitors for all colonic epithelial cells, are some of the most actively dividing stem cells in the body. The entire colonic epithelium is estimated to be replaced weekly, with an average cell cycle length of a day[30,31]. To control the vast

numbers of bacteria, specific colonic cells secrete a cocktail of anti-bacterial peptides[32,33]. Taken together, the colon combines rapid cell growth, mucus secretion, and antimicrobial peptide secretion to maintain a single cell layer capable of water absorption, bacterial control, and the passing of feces[34].

2.2.2 Composition of the Human Colon

The colon is comprised of several concentric layers of various tissues whose coordinated activity results in proper colon function. The layers of the colon include: the mucosa, the submucosa, the muscularis propria, and the serosa (Figure 4).

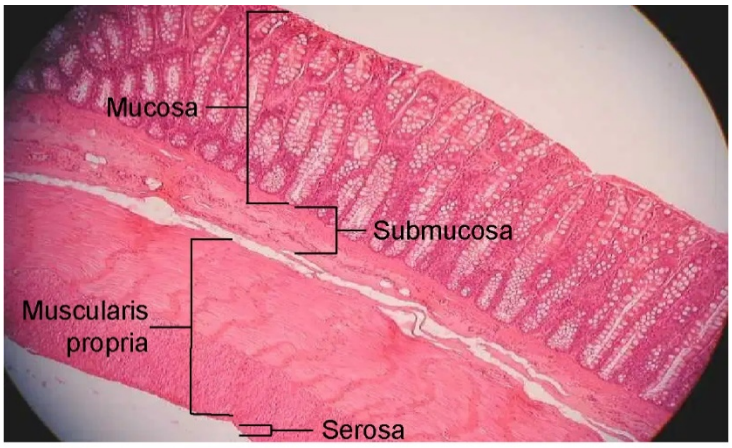

Figure 4 | Histology of the normal colon. The mucosa lies adjacent to the intestinal lumen and contains the epithelial cell layer. The submucosa connects the mucosa to the muscularis propria, and the serosa forms the outer layer of the tissue.

The inner most layer, the **mucosa,** consists of three sublayers: the epithelium, the lamina propria, and the muscularis mucosae[27]. The epithelium is a single cell layer lining the interior of the lumen responsible for mediating water absorption, mucus secretion, and bactericidal peptide secretion. It also forms a physical barrier for the body. The epithelium forms multiple invaginations, referred to as crypts, which are comprised of several cell types, each performing a specific function (Figure 5).

The cell types of the colon and their functions are:

Enterocytes: Physical barrier, water, and ion absorption

Goblet cells: Secrete mucus, transfer microbial antigens to dendritic cells

Paneth cells: Secrete antimicrobial peptides, maintain intestinal stem cells[35]

Tuft cells: Regulate the activity of lymphoid cells

Intestinal stem cells (CBC): Generate the cell types of the colon

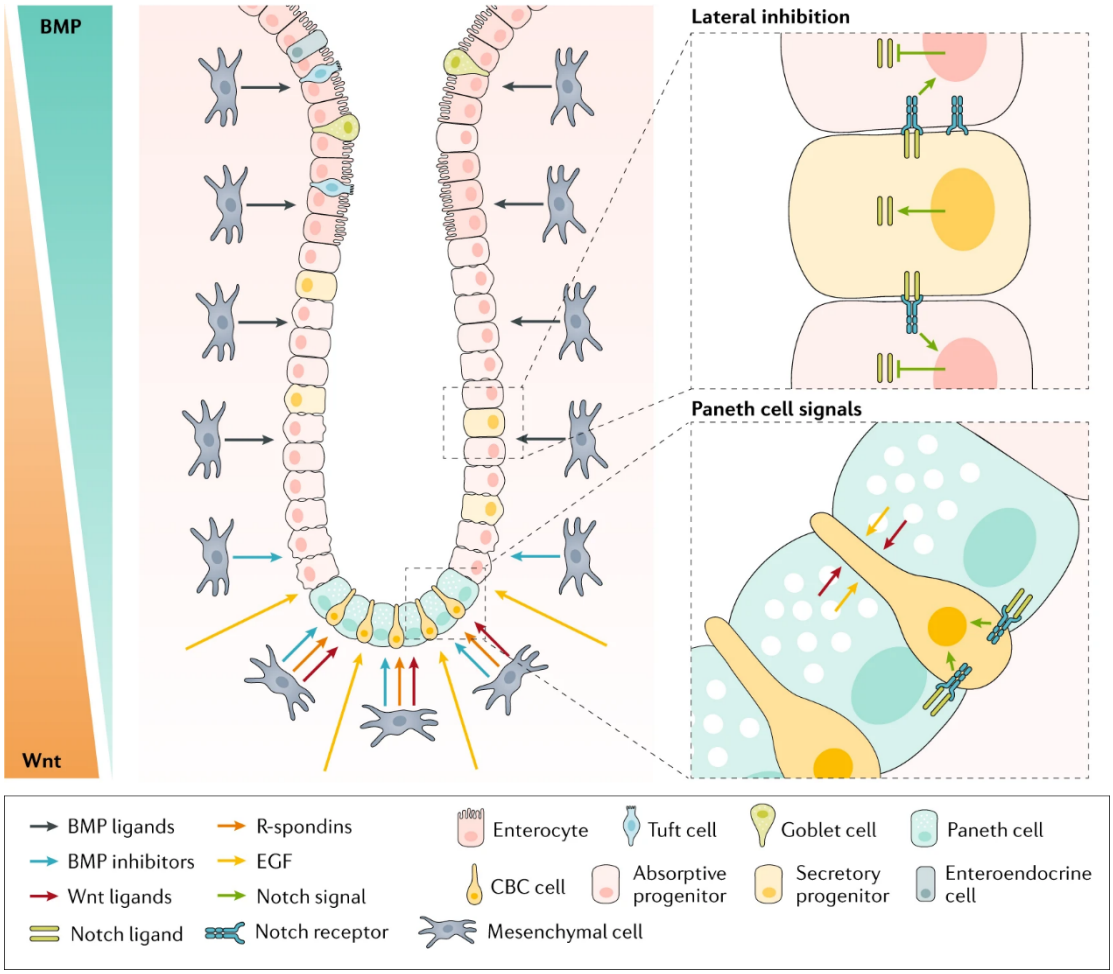

Figure 5 | Epithelial structure of a colonic crypt. Reproduced with permission from Gehart and Clevers, 2019.

The cells of the intestine are organized in a hierarchy along the axis of the intestinal crypts, which is crucial to their development and function[36]. The stem cells, also referred to as crypt columnar base (CBC) cells, reside at the base of the colonic crypt, interspersed with Paneth cells[37]. The stem cells are maintained by growth stimulatory signals from the surrounding Paneth cells as well as the mesenchyme. As the stem cells divide, the progeny cells are pushed outside of the stem cell niche and begin to differentiate. Asymmetrical cell division does not appear to take place in the intestinal stem cells, therefore exclusion of the stem cells from the stem cell niche appears to be stochastic[31,38]. These excluded cells continue to divide as they ascend the crypt and are referred to as transit amplifying cells. As the cells approach the luminal surface (the top of the crypt), they fully differentiate into the various intestinal cell types mentioned previously, and stop proliferating[39]. Replicating cells are found in the bottom 75% of the crypt, while replication is absent in the upper 25% of the crypt[40]. Upon reaching the luminal surface, the cells are either removed by the movement of the feces, or are exfoliated into the lumen via apoptosis, with another cell quicky taking its place. A delicate balance between cell proliferation and death maintains proper colonic epithelial architecture with the entire crypt being renewed weekly (~6 days)[30,41].

The epithelium is supported by the lamina propria, which is an interstitial tissue comprised of an extensive lymphatic network[27]. The lamina propria is comprised of collagen and a variety of cell types including fibroblasts, lymphocytes, eosinophils, macrophages, and mast cells. The function of the lamina propria is to structurally support and defend the epithelium[42]. Microbial pathogens which have breached the epithelium are met with a dense collagenous matrix housing a plethora of lymphoid

cells. The third, and final, layer of the mucosa is the muscularis mucosae, which consists of smooth muscle fibers. Contraction of the muscularis mucosae may facilitate the movement of secreted compounds into the lumen of the intestine.

The **submucosa** is a dense layer of connective tissue, similar in composition to the lamina propria, underneath the mucosa containing collagenous fibers, blood vessels, fibroblasts, macrophages, mast cells, and nerves[27]. The submucosa is innervated by two neural plexuses, the Meissner plexus, and the Auerbach plexus. The submucosa can stretch with increased capacity, however, maintains the cylindrical shape of the colon. Ganglia from the submucosa extend into the muscularis mucosae of the mucosa. The submucosa is vascularized by the arterioles, venules, and lymphatic vessels. Invasion of the submucosa by neoplastic cells of the colonic epithelium therefore marks the transition to malignant disease due to access to the vasculature[43].

The **muscularis propria**, also called the muscularis externa, lies outside the submucosa and is comprised of two layers of smooth muscle. The inner layer, closer to the submucosa, is arranged in rings which encompass the inner layers of the colon. The outer, longitudinal layer lies perpendicularly to the inner layer. These muscles are responsible for peristalsis and are controlled by the myenteric plexus, which lies between the two muscle layers. Contractions are initiated by the interstitial cells of Cajal. The gut has intrinsic peristaltic activity due to its self-contained nervous system.

The **serosa** is the outermost layer of the colon and is comprised of several layers of connective tissue covered by a squamous epithelium, referred to as the mesothelium, which reduces friction between the colon and the surrounding organs during digestion.

2.2.3 Aberrant Crypt Foci

Aberrant Crypt Foci (ACF) are colorectal crypts in which the epithelial cells have begun to proliferate abnormally. They represent the earliest identified precursors to colonic polyps and colon cancer, and can be found in ~60% of adults 50-59 years old[44]. ACF are divided into two main types, nondysplastic (hyperplasia) and dysplastic (dysplasia), based on their morphology[45] (Figure 6). Nondysplastic ACF presents either as a single crypt, or multiple neighboring crypts, with an enlarged, thickened morphology[46-49]. The luminal surface may be serrated and extend slightly beyond the surface of the normal mucosa, however cells with positive staining for proliferation markers remain in the middle and lower regions of the crypt(s)[46-48]. Cell nuclei may be enlarged and mucin, indicative of differentiated Goblet cells, may be partially depleted[49]. Dysplastic ACF presents with enlarged and elongated crypts containing cells with enlarged nuclei, loss of nuclear polarity, and nuclear hyperchromatism. Goblet cells are significantly decreased in number, and positive staining for proliferative markers is observed throughout the crypt, including the top[48,49]. While all ACF are considered benign, dysplastic ACF have the greatest likelihood of developing into colonic polyps[48]. Though the less threatening nondysplastic ACF are more commonly observed (90%) than the potentially malignant dysplastic ACF (10%)[47-50].

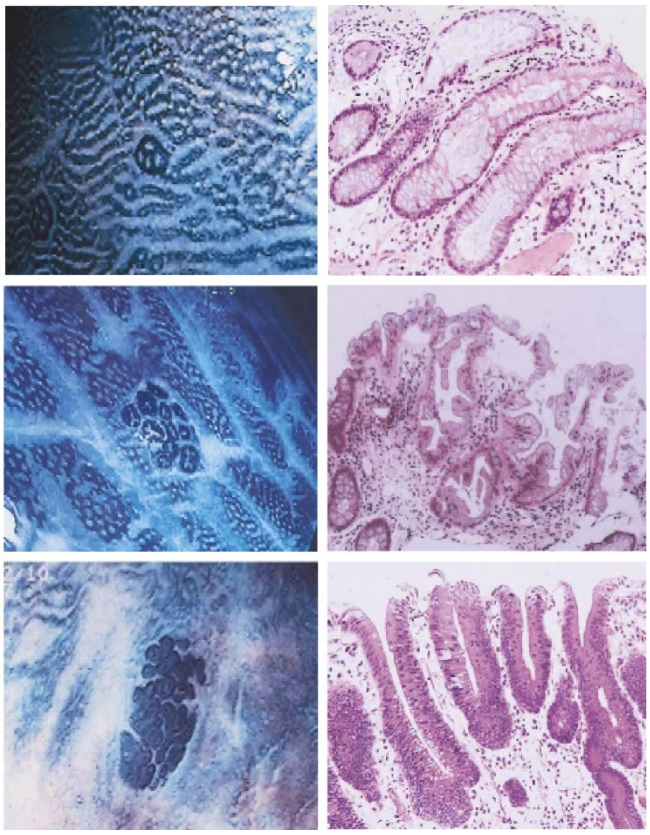

Figure 6 | Morphology of aberrant crypt foci. In the top left image, methylene blue staining was used to visualize a small aberrant focus. The aberrant crypts, seen in the center of the image, stain more darkly and have a thicker epithelial lining than normal crypts. Histologically (top right), there is a slight elongation and enlargement of the ducts, which was visualized using hematoxylin and eosin (H&E) staining. In the middle left figure, a hyperplastic ACF shows the typical slit-shaped lumen in methylene blue staining, and a serrated morphology with H&E. In the bottom left figure, a dysplastic focus with a deformed and slightly raised shape is seen. The lumens of the crypts are compressed. This large ACF shows the histological signs of dysplasia, seen bottom right, including enlarged and enlongated crypts, loss of nuclear polarity as well as many hyperchromatic nuclei. Reproduced with permission from Takayama et al., 1998, Copyright Massachusetts Medical Society.

The prevalence of ACF, in general, increases with age with ~10% of normal patients under the age of 40 presenting an ACF, which rises to 66% in patients 60 to 69 years of age[44,48]. However, patients diagnosed with adenomas or carcinomas, have an increased ACF burden at all ages, suggesting a link between the three[44,51]. Another similarity is the anatomical distribution of ACF, colonic polyps, and colon cancers, with all three occurring primarily in the left colon (descending and sigmoid)[46-48]. Microscopic analysis has demonstrated that dysplastic ACF are morphologically similar to polyps[47,49,50]. At the molecular level, the most common genetic alteration found in colon cancer, mutation of *APC*, is also present in some dysplastic ACF. Additionally, patients with FAP, an inherited disease which bears significant colorectal cancer risk, have an increased burden of dysplastic ACF, suggesting a common growth mechanism.

2.2.4 Colonic Polyps

Epithelial polyps are localized, well-demarcated tumors which typically project above the surface of the surrounding mucosa, however flat and depressed neoplasms also occur[52]. The abnormal tissue architecture first observed in ACF is exacerbated and epithelial polyps are larger than ACF. They are present in ~45% of adults over 50 years old and are divided into two main types, non-neoplastic and neoplastic, based on their morphology[53]. Non-neoplastic polyps, also referred to as hyperplastic polyps, have a serrated morphology in the top-half of their elongated crypts[54]. No dysplasia is evident, and the risk of malignant progression is minute[54-56]. Neoplastic polyps, also referred to as adenomas, present dysplastic growth and may refer to: traditionally serrated polyps, adenomatous polyps, or mixed polyps[54]. Sessile serrated polyps are also referred to as neoplastic polyps or adenomas, though they are not dysplastic. Instead, they contain L- or T-shaped crypts in which the proliferative zone of the crypt has shifted up to the wall of the crypt[54]. The neoplastic polyps are characterized based on morphology, with the sessile serrated and traditionally serrated polyps, as their names imply, presenting a serrated morphology[54]. Adenomatous polyps exhibit a smooth, lobulated morphology, while mixed polyps present a combination of the two[56]. The majority (~80%) of the serrated polyps (hyperplastic, sessile serrated, and traditionally serrated) are hyperplastic, while ~15% are sessile serrated polyps[54]. Traditionally serrated polyps, which are the rarest, account for ~5%[54]. The majority (~90%) of polyps are non-neoplastic, while the remaining ~10% are neoplastic[43].

The histology of adenomatous polyps is characterized as tubular, tubulovillous, or villous, which aids in classifying the stage of the adenoma[56] (Figure 7). Tubular adenomas consist of closely packed epithelial tubules, separated by lamina propria, which branch horizontally to the muscularis mucosae[56]. The tubules can have a regular, well-differentiated structure, or may be highly branched. Villous adenomas consist of finger-like processes, comprised of a lamina propria core, covered by epithelial cells growing vertically towards the lumen[56]. The epithelium rests on the muscularis mucosae when not supported by the processes of the lamina propria. Tubulovillous adenomas contain a combination of both tubule and villous morphologies. Adenomas with more than 75% villous features are classified as villous adenomas, while those with 25% to 75% villous features are tubulovillous.

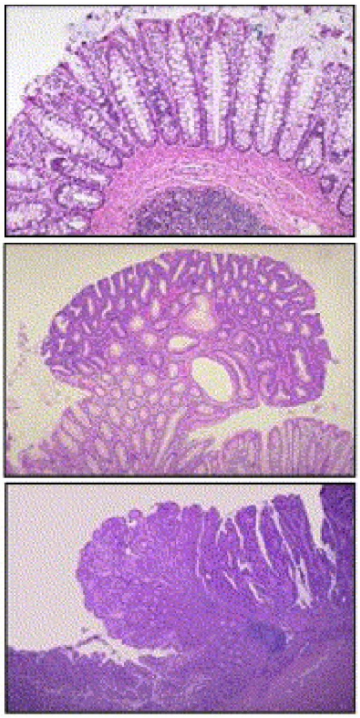

Figure 7 | Histology of colon tumorigenesis. The topmost image shows the normal colonic epithelium with well-organized crypts. The center image shows the more unruly organization of a tubular adenoma. The bottommost image reveals the chaotic histology of a tubulovillous carcinoma. Reproduced with permission from Cardoso et al., 2007.

Adenomas which contain at least 25% villous features, or high-grade dysplasia, are considered advanced, given their higher malignant potential[57,58]. Therefore, villous and tubulovillous adenomas are considered advanced. The majority (~75%) of adenomas are tubular, while ~15% are tubulovillous, and ~10% are truly villous[56,57]. While the carcinoma stage has yet to be described, Figure 7 demonstrates the increasing loss of tissue organization observed during the progression to colon cancer.

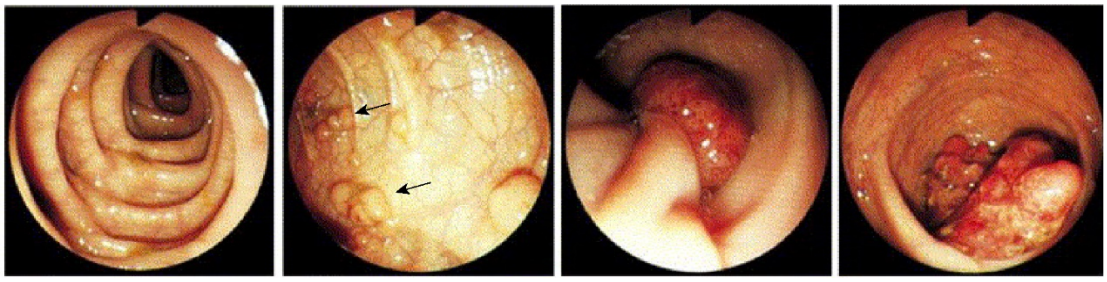

Figure 8 | Phenotypes of colon tumorigenesis. The leftmost image is indicative of a normal colon, with a smooth luminal lining. The left-center image shows small polyps (some shown by arrows). The right-center image shows a large adenomatous polyp, while the rightmost image shows a colon carcinoma. Reproduced with permission from Cardoso et al., 2007.

Size is also an important marker for classifying adenomas, and polyps are grouped into three different categories: diminutive (1 to 5 mm in diameter), small (6 to 9 mm), or large (>10 mm)[57-59] (Figure 8). Large adenomas are considered advanced, regardless of histology. Approximately 60% of polyps are diminutive or small, while ~40% are large[58]. Additionally, as polyps grow, their structural architecture changes. The majority (~75%) of tubular adenomas are small polyps, ~50% of tubulovillous polyps range from 10-20mm, and ~60% of villous polyps are 20mm or larger[58]. However, ~15% of villous polyps are 9mm or less, demonstrating that even small lesions have the potential to progress towards malignancy.

Neoplastic colonic polyps are considered the successor of ACF and the predecessor of colon cancer since patients diagnosed with polyps or carcinomas, have an increased ACF burden at all ages and these three growths occur primarily within the same anatomical location, the sigmoid colon[44,57]. Additionally, the occurrence of ACF, adenomas, and colon cancer all increase with age[44,60]. Morphologically, polyps resemble dysplastic ACF, though with a more severely disorganized histology[56]. However, some dysplastic ACF have been referred to as microadenomas, reflecting their morphological similarities[47,50,61]. In regards to malignant progression, remnants of

adenomas have been found in carcinomas, with higher incidence found in earlier stage carcinomas (those which have not yet breached the submucosa) and microscopic foci of carcinoma, have been found in adenomas[55,58,62,63]. Furthermore, the most common genetic alterations occurring in colon cancer, some of which are observed in ACF, are also present in colonic polyps[64,65]. Colonic polyps therefore represent the intermediary stage in colon tumorigenesis as they demonstrate a similar yet more severe histology to dysplastic ACF, but nevertheless exhibit more organized tissue architecture than observed in colon cancer. The shared characteristic mutations of colon cancer and progressive deterioration of tissue architecture support a sequence from ACF-adenoma-carcinoma. This sequence takes from 5-30 years to develop[58].

2.2.5 Colon Cancer

A colonic tumor is classified as a carcinoma when it displays high-grade dysplasia[66]. Its growth pattern ranges from polyp-like masses of tissue to flat lesions which are often inflamed[27,66]. The polyp-like lesions protrude into the lumen, and as they are usually asymptomatic, can grow to large sizes (>20mm). The flat lesions are ulcerated with raised edges surrounding the luminal side of the tumor mass. A 2-tiered grading system is used to describe carcinoma histology[67]. Tumors are considered low-grade when 50% or more of the tumor is gland forming, while a high-grade tumor contains less than 50% gland forming tissue[67]. Generally, low-grade is associated with higher patient survival, likely due to retention of differentiated cell function[67-69].

Colon carcinoma is also pathologically staged, according to the TNM (Tumor, Node, Metastasis) Staging System of the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC)[15,67,70]. The ‘T’ refers to the first

resection of the primary tumor. The 'N' refers to the status of the regional lymph nodes, and the 'M' refers to distant metastatic disease. TNM code definitions are given in Table 1 and staging information in Table 2. For example, T1 refers to a Stage I colon carcinoma that has invaded the submucosa, which has not yet metastasized. Tumor TNM staging is the most useful tool in determining survival[69,71,72].

Code	Phenotype
Primary Tumor (T)	
T0	No evidence of primary tumor
Tis	Carcinoma in situ (intraepithelial or intramucosal carcinoma)
T1	Tumor invades the submucosa
T2	Tumor invades the muscularis propria
T3	Tumor invades through the muscularis propria into the subserosa or into the non-peritonealized pericolic or perirectal tissues
T4	Tumor directly invades other organs or structures (T4a) or perforates the visceral peritoneum (T4b)
Regional Lymph Nodes (N)	
N0	No regional lymph node metastasis
N1	Metastasis in 1-3 lymph nodes
N2	Metastasis in 4 or more lymph nodes
Distant Metastasis (M)	
M0	No distant metastasis
M1	Distant metastasis

Table 1 | TNM code descriptions (left). Reproduced with permission from Compton et al., 1999.

Stage	TNM Codes
Stage 0	Tis N0 M0
Stage I	T1 N0 M0
	T2 N0 M0
Stage II	T3 N0 M0
	T4 N0 M0
Stage III	Any T N1 M0
	Any T N2 M0
Stage IV	Any T Any N M1

Table 2 | Colon cancer staging according to TNM codes (above). Reproduced with permission from Compton et al., 1999.

The majority of colon cancers (95%) are adenocarcinomas, which are cancers originating in gland-forming epithelium. A fraction of adenocarcinomas (~10%) are mucinous, while less frequent types include signet-ring cell adenocarcinoma (1%), medullary-type adenocarcinoma (0.1%), and squamous carcinoma (<0.01%)[27,66].

Adenocarcinoma of the colon is the most frequent carcinoma observed. The majority of adenocarcinomas are low-grade at diagnosis, though ~40% present with metastasis to regional lymph nodes, and ~20% have metastasized to the liver at diagnosis[27,43,66]. The 5-year survival rate of patients without metastasis is approximately 80%, which decreases to ~50% in patients with a regional lymph node metastasis[15,72]. Survival drops precipitously for patients with distant metastatic disease to ~3%. Metastasis is therefore the greatest determinant of survival. Carcinoma size, which contrasts with the importance of size at the adenoma stage, does not influence survival[71]. Metastasis becomes possible once the cancerous cells have breached the muscularis mucosae and have accessed the submucosa, where they can enter the lymph and blood vessels[43]. Tumors limited to the mucosa have not been observed to metastasize[43].

The most common sites of metastasis are determined by the vasculature of the body. Adenocarcinomas of the right colon first metastasize to the ileocaecal, right colic, and middle colic lymph nodes, and subsequently, to the superior mesenteric nodes. Lymph from the left colon is drained to the left colic, sigmoidal, and lower mesenteric lymph nodes. Hematogenous metastasis usually occurs first in the liver, since the large majority of venous blood leaves the colon through the portal system, which flows towards the liver. The lungs are the second most commonly colonized site[15,27].

Mucinous Adenocarcinoma is an infrequent (~10%) variant of adenocarcinoma wherein 50%, or more, of the tumor volume is comprised of extracellular mucin[73,74]. Prognosis of patients with mucinous adenocarcinoma is controversial, though many occur in patients with hereditary nonpolyposis colorectal cancer (Lynch Syndrome), and thus are commonly microsatellite instable – high (MSI-H) tumors, which typically behave in a low-grade fashion[75,76]. However, mucinous adenocarcinomas may also be microsatellite stable (MSS) and therefore behave more aggressively.

Signet Ring Cell Adenocarcinoma is rare in the colorectum and accounts for ~1% of all colorectal carcinomas[77]. These carcinomas are defined by the presence of a large mucin-filled vacuole, which pushes the nucleus to the periphery of the cell. However, more than >50% of the tumor cells must have a mucin-filled vacuole to classify the tumor as a signet ring cell adenocarcinoma[78]. These tumors are poorly differentiated and carry a worse outcome than conventional adenocarcinoma[66,79]. Metastasis to the peritoneum is frequent, though hematogenic metastasis to the liver or lungs is rare[78,80]. However, some may be MSI-H, and act in a low-grade fashion[66].

Medullary carcinoma is extremely rare and accounts for <0.1% of all colorectal cancers[81]. This tumor is characterized by sheets of neoplastic, epithelioid cells with large, vesicular nuclei, prominent nucleoli, and intraepithelial lymphocytic infiltrate[66,82]. This type of cancer is usually MSI-H and therefore has a favorable prognosis[82].

2.2.6 Colon Cancer Development

Colon cancer develops in a stepwise progression with growth becoming increasingly aberrant as cells progress from ACF to adenoma to carcinoma (Figure 9). This process, which can take decades, appears to progress through two mutually exclusive pathways. The chromosome instability (CIN) pathway, which accounts for ~85% of colon cancers, and the microsatellite instability (MIN) pathway (~15%)[83,84].

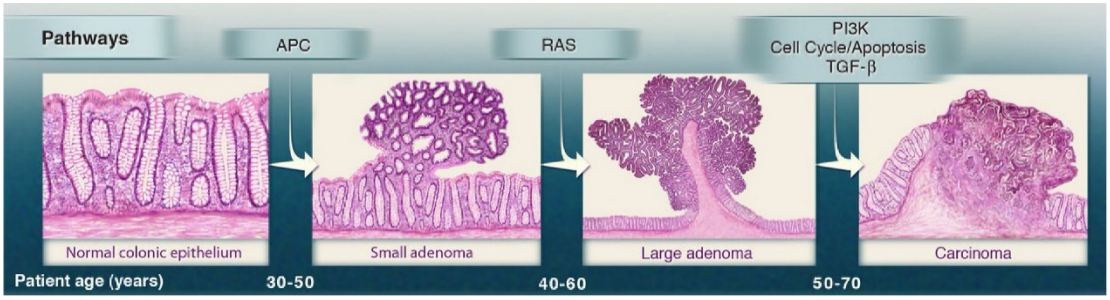

Figure 9 | The path to colon cancer. This diagram demonstrates the most common progression from normal epithelium to carcinoma. Reproduced with permission from Vogelstein et al., 2013.

The **CIN** pathway is initiated by a genetic perturbation which activates the WNT signaling pathway, most commonly via mutation or loss of *APC*, a key negative regulator of WNT signaling[64,85,86]. APC plays a crucial role in colorectal tumorigenesis as inheriting an *APC* mutation, or *APC* loss, causes familial adenomatous polyposis (FAP), a condition which predisposes affected individuals to colorectal cancer[87-89]. *APC* mutation results in the growth of a dysplastic ACF, though prevalence of *APC* mutations in ACF is generally low[64,90]. The low rate (~4%) of *APC* mutation in ACF may explain the fact that most ACF do not progress into an adenoma. In support of this, the majority of adenomatous polyps (~70%) harbor an *APC* mutation, indicating the selection of mutated *APC*[64,85]. Chromosomal instability, or aneuploidy, which is the characteristic feature of the CIN pathway, also occurs early in colorectal tumorigenesis. The acquisition and maintenance of specific chromosomal

aneuploidies, such as the gain of chromosome 7, suggests selection, and supports aneuploidy as a cause, as opposed to consequence, of colorectal cancer[91,92]. Mutation of *APC* combined with aneuploidy results in a small adenoma. Additional mutations are required for the progression from adenoma to carcinoma[83,93-95]. *KRAS* is frequently mutated in colorectal adenocarcinomas, and mutations are found in ~25% of both dysplastic and non-dysplastic polyps[64]. Given the relatively equal distribution of *KRAS* mutations between dysplastic and non-dysplastic polyps, this suggests that *APC*, and not *KRAS*, is the driving force behind dysplasia. Instead, the mutation of *KRAS* in an *APC* mutated adenoma results in an increased growth rate and clonal expansion, advancing a small adenoma into a larger one[93,96]. The *APC* and *KRAS* mutations therefore function together to promote tumor growth. Inactivating mutations in *SMAD4* and *TP53*, or loss of these genes due to chromosomal aneuploidy, complete the malignant transformation and advance the large adenoma into a carcinoma[83,93-95]. Loss of 17p (*TP53*) is infrequently observed in adenomas, yet is present in more than ~75% of carcinomas, identifying it as a late event in the CIN pathway[93,96,97].

The **MIN** pathway is initiated by mutation of the *KRAS* or *BRAF* proto-oncogenes, which leads to the development of hyperplastic ACF and hyperplastic polyps[90,98-101]. *KRAS* mutations are found in ~65% of ACF[98,100]. However, the prevalence of *KRAS* mutation decreases to ~20% in hyperplastic polyps, while BRAF mutations are found in ~70% of hyperplastic polyps[102]. Methylation-induced silencing of various tumor suppressor or DNA repair genes, referred to as a CpG island methylator phenotype (CIMP) then promotes the progression to a serrated adenoma[102-105]. The prevalence of *KRAS* and *BRAF* mutations in serrated adenomas was similar to that observed in

hyperplastic polyps (~30% and 60%, respectively)[102]. Mutations in *KRAS* or *BRAF* appear to be mutually exclusive[102]. The final step is the silencing or loss of *MLH1*, which leads to microsatellite instability, the characteristic feature of this pathway. Loss of *MLH1* is associated with high-grade dysplasia[104].

2.3 Colon Cancer: Etiology

2.3.1 WNT Signaling

WNT signaling is an ancient pathway found in all three major metazoan clades (deuterostomes, ecdysozoans, and lophotrochozoans)[106]. Across the animal kingdom, WNT signaling is responsible for regulating cell growth and spatial orientation in developing embryos and maintaining stem cells in adult tissues[107]. WNT signaling is therefore crucial throughout the lifespan of the organism.

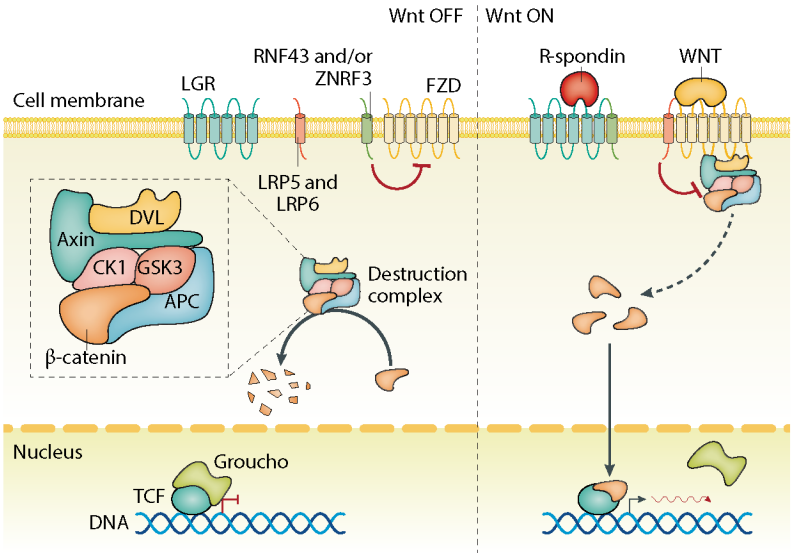

Figure 10 | Diagram of the WNT Signaling Pathway in the OFF (left) and ON (right) states. Reproduced with permission from Gehart and Clevers, 2019.

Canonical WNT signaling is initiated by a WNT ligand, which is a small (40kDa), cysteine rich protein, called a WNT[108] (Figure 10). Humans have 19 WNTs, with limited functional overlap, as distinct phenotypes result from the genomic deletion of various WNTs[108,109]. As part of their maturation, WNTs are lipid modified by the addition of a palmitoleic acid, which is necessary for secretion and receptor binding[110,111]. The lipid group is attached to the WNT (Ser-209 on WNT3A) by Porcupine, an O-acyltransferase, in the endoplasmic recticulum[111,112]. The lipidated WNT is transferred

to the Golgi where it is bound by Wntless, a transmembrane protein, which is required for ferrying WNTs to the plasma membrane for presentation or secretion[113-115]. While most WNT interactions likely occur via cell-to-cell contact, it remains unclear whether WNTs may also act at longer ranges, like typical morphogens[107,116]. It has been shown that WNTs may traverse to their target cell on the surface of secretory vesicles in an exposed position for binding[117,118]. However, diffusion of a lone WNT through the extracellular environment is unlikely due to palmitoylation-induced hydrophobicity[110].

WNT binds two receptors, FZD and LRP5/6, at the target cell plasma membrane to form a complex[119,120]. FZD receptors are comprised of a 7-transmembrane domain and a cysteine-rich domain (CRD), which lies in the extracellular environment[121]. WNT ligands bind the CRD domain with nM affinity in a pincher-like conformation formed by the palmitoyl-group and the WNT C-terminus[119,120,122]. Humans have 10 FZD receptors which bind promiscuously with multiple, but not all, WNTs[107]. The downstream consequences of all WNT-FZD binding interactions are still under investigation, though several WNTs initiate the canonical WNT signaling pathway while others are involved in non-canonical WNT signaling. The remaining coreceptor, either LRP5 or LRP6, is a single-span transmembrane protein with extracellular Epidermal Growth Factor (EGF) repeats and Low-Density Lipoprotein Related (LDLR) repeats linked via a single-span transmembrane domain to five intracellular Pro-Pro-Pro-Ser-Pro (PPPSP) motifs[123,124]. Association of FZD with LRP5/6, caused by WNT binding, results in the phosphorylation of the PPPSP motifs of LRP5/6, generating a docking site for the cytoplasmic scaffolding protein AXIN on the FZD/LRP complex at the plasma membrane[123-125]. In the cytoplasm, AXIN organizes a complex comprised of APC, CKI,

GSK3, and β-catenin, referred to as the destruction complex. Relocation of AXIN from the cytoplasm to the FZD/LRP complex, which requires the protein DVL, relocates the destruction complex to FZD/LRP[126,127]. Once relocated, the destruction complex is unable to facilitate the proteasomal degradation of β-catenin[128]. This results in the accumulation of phosphorylated β-catenin in the destruction complex, now sequestered to the plasma membrane, as well as the accumulation of β-catenin in the cytoplasm[128]. Once sufficient cytoplasmic levels have been reached, β-catenin migrates to the nucleus where it binds the TCF/LEF family of transcription factors, which mediate canonical WNT signaling target gene expression[129,130].

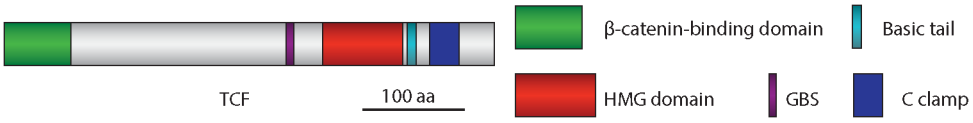

Figure 11 | Typical structure of the TCF/LEF family of transcription factors. Reproduced with permission from Cadigan and Waterman, 2012.

Humans have four TCF/LEF transcription factors, TCF1, TCF3, TCF4, and LEF1, which are encoded by *TCF7*, *TCF7L1*, *TCF7L2*, and *LEF1*, respectively[131-135]. These transcription factors are comprised of a highly conserved HMG box, a basic tail, a β-catenin binding domain, a TLE binding domain, a CtBP binding domain, and a C-clamp[136] (Figure 11). These transcription factors can not activate or repress transcription on their own, as they do not contain the appropriate domains[129,137]. Instead, their function is determined by their binding partner(s). The HMG box confers DNA sequence specific binding, while the basic tail binds the DNA backbone and functions as a nuclear localization signal[131-133,138]. The TCF/LEF HMG box recognizes a consensus sequence of 5' – SCTTTGATS – 3' and binds the minor groove of DNA, which results in a sharp bend (130°) in the DNA[131,139-141]. The β-catenin binding domain

lies at the N-terminus of the TCF/LEFs and, when bound to β-catenin, converts the TCF/LEF factors into potent transcriptional activators[129,130]. Deletion of this domain results in truncated, dominant negative proteins which no longer bind β-catenin and instead, suppress WNT signaling activity[129]. Suppression of WNT signaling activity occurs through a variety of proteins, however the best studied are the Transducin-Like-Enhancer of Split (TLE) repressors[136,142,143]. The TLE factors bind at multiple sites along the TCF/LEF proteins, specifically the central portion, as well as the HMG domain[144,145]. However, β-catenin binding displaces the inhibitory TLE factors, converting the TCF/LEF factors into transcriptional activators[144]. Another group of inhibitory co-repressors are the CtBP proteins, which bind to the C-terminus of the TCF/LEFs[146,147]. The C-clamp is the final domain of the TCF/LEF factors and functions as a second DNA binding domain. It is called a C-clamp due to four conserved cysteine residues required for DNA binding[141,148,149]. The C-clamp binds a GC-rich helper sequence, which may facilitate more stable binding of the TCF/LEF factors to the DNA at specific sites where the GC-rich motif is appropriately spaced with the HMG binding sequence[148,149]. However, not every TCF/LEF protein has a C-clamp[136,148]. The downstream consequences of WNT signaling therefore depend upon the TCF/LEF factors present, the relative abundance of their co-activators and co-repressors, as well as the sequence and spacing of the gene regulatory elements at the locus.

TCF1 plays a role in the development and differentiation of the thymocytes[150]. Loss of TCF1 in mice results in a relatively mild phenotype as animals are viable and fertile, however display a block in T-lymphocyte differentiation[150]. The mild phenotype is due to functional redundancy with LEF1, which also facilitates development of the immune system[132,133]. TCF1 regulates transcription by two mechanisms: binding with β-catenin to activate target gene expression, or by binding to the TCFα enhancer and bending the DNA such that two separate motifs along the genome become closely oriented in three-dimensional space[131,134]. However, different isoforms of TCF1, which lack the β-catenin binding domain, are found in mammary and intestinal tissue[151]. These isoforms function as repressors since they retain the TLE binding domain[151]. Adult animals lacking TCF1 develop higher numbers of mammary gland and intestinal adenomas, indicating a tumor suppressive role for this transcription factor[151]. *TCF7* contains the C-clamp, however lacks the CtBP binding domain.

TCF3 is one of the earliest expressed WNT transcription factors and mouse embryos lacking *TCF7L1* die around embryonic day 9.5 (E9.5)[152]. In contrast to the generally activating roles of other WNT transcription factors, TCF3 appears to be largely repressive[146,153]. In adults, *TCF7L1* is highly produced in several organs including the cervix, breast, colon, and adipose tissues[154]. *TCF7L1* does not contain the C-clamp, however does include binding sites for the transcriptional repressor, CtBP[154].

TCF4 is the best studied of the TCF/LEF factors. It is necessary for development of the gut in embryonic mice as well as the maintenance of the colonic epithelium in adult mice, establishing TCF4 as a crucial transcription factor in the colon[155,156]. Perturbation of *TCF7L2* results in loss of the proliferative compartments of the intestinal epithelium, demonstrating that TCF4 is necessary for maintaining the stem cells[155,156]. Activating WNT pathway mutations lead to increased levels of nuclear β-catenin, which results in the activation of WNT target genes, such as *MYC* and *CCND1*, via βcat/TCF4 complexes[157,158]. *TCF7L2* contains both the C-clamp and the CtBP binding domains.

LEF1 is widely expressed during mouse embryonic development and mice lacking *LEF1* die one to two weeks after birth due to skin, hair, teeth, and mammary gland defects[159]. While LEF1 also plays a role in lymphocyte development, no apparent alteration in the lymphocyte population was observed upon *LEF1* perturbation, likely due to the action of TCF1, which is also expressed in lymphocytes[131,134,159]. LEF1 activates target gene expression by two mechanisms: binding with β-catenin to activate gene transcription, or by binding to the TCRα enhancer and, via bending of the DNA, aligning the binding motifs of other transcription factors[132,133]. *LEF1* does not contain the C-clamp domain or the CtBP binding domains[136,146].

When the FZD and LRP5/6 coreceptors are not bound by a WNT ligand, AXIN is not sequestered to LRP5/6. Instead, AXIN remains in the cytoplasm, bound to the members of the destruction complex: APC, β-catenin, CKIα, and GSK3. CKIα phosphorylates β-catenin on Ser45, which is followed by the sequential phosphorylation of β-catenin on residues Thr41, Ser37, and Ser33 by GSK3[160]. Phosphorylation of β-catenin generates a motif recognized by β-TrCP, an F-box-containing ubiquitin ligase, leading to the ubiquitination and proteasomal degradation of β-catenin[161,162]. In the absence of a WNT ligand, β-catenin is rapidly degraded and does not accumulate in the cytoplasm. The TCF/LEF transcription factors remain bound to TLE factors in the nucleus, repressing target gene expression.

2.3.2 RAS Signaling

RAS signaling is crucial for mediating cellular growth, differentiation, and survival[163]. Members of the RAS signaling pathway, most notably *KRAS*, are frequently mutated in cancer[164]. Approximately 50% of all colon cancers harbor a mutation in *KRAS*[164]. The key event in the activation of RAS signaling is the exchange of GDP with GTP on RAS, which is facilitated by Guanine Nucleotide Exchange Factors (GEFs)[165]. The inactivation of RAS signaling is mediated by GTPase-Activating Proteins (GAPs), which induce GTP hydrolysis and a return to the GDP-bound state of RAS[166,167]. When activated (bound by GTP) RAS activates the protein serine/threonine kinase RAF (c-RAF1, BRAF, or ARAF)[168]. Then, activated RAF phosphorylates, and by doing so activates, the mitogen-activated protein kinase kinases 1 and 2 (MEK1 and MEK2), though MEK1 and MEK2 may also be activated independently of RAF[169]. These kinases then activate the mitogen-activated protein kinases (MAPKs) ERK1 and ERK2[170]. The MAPKs can be transported to the nucleus following activation, resulting in the activation of both cytosolic and nuclear substrates. ERK phosphorylates a variety of transcription factors including the ETS transcription factors, as well as c-JUN[168,171]. The stimulation of these transcription factors results in the expression of key cell-cycle regulatory proteins which promote proliferation and survival. Mutation of *KRAS* in colon cancer holds an additional benefit, as c-JUN, which is activated by *KRAS*, interacts with β-catenin and TCF4, thereby stabilizing WNT signaling activity.

2.3.3 TGFβ Signaling

TGFβ signaling influences a diverse array of cellular processes throughout the lifespan of the organism including proliferation, differentiation, morphogenesis, and regeneration[172]. In the colonic epithelium, finely regulated gradients of TGFβ signaling alongside WNT signaling mediate differentiation of the intestinal stem cells and in colon cancer, mutations influencing TGFβ signaling activity are typically one of the last genetic perturbations to occur before the transition to colon carcinoma[34,94]. TGFβ signaling is initiated by the binding of a ligand from either the TGFβ-activin-Nodal or the BMP subfamilies[173]. Ligand binding results in the formation of a receptor complex consisting of two Type I (signal-propagating) and two Type II (activating) components, which are both serine/threonine kinases[174]. Complex formation results in phosphorylation of the Type I receptor by the Type II receptor, which allows the Type I receptor to activate the receptor-associating SMAD (R-SMAD) proteins[174]. SMAD1, SMAD5, and SMAD8 are phosphorylated by the BMP-binding receptors, while SMAD2 and SMAD3 are phosphorylated by the TGFβ-, activin-, and Nodal-binding receptors[173]. Activated R-SMADs then bind with SMAD4, which acts as a promiscuous partner of all R-SMADs. Trimers of two R-SMADs and one SMAD4 are considered the principal functional units of TGFβ-mediated transcriptional activity[173]. The R-SMAD/SMAD4 trimers are then imported into the nucleus, where they associate with other DNA-binding transcription factors to facilitate target gene expression[175,176].

Chapter 3: Methods

3.1 Cell Culture Techniques

3.1.1 Cell Passaging

The COLO201, DLD1, HCT116, HT29, LoVo, LS174T, SW480, and SW620 colon cancer cell lines were purchased from the American Type Culture Collection (ATCC). COLO201, DLD1, SW480, and SW620 were cultured in RPMI-1640 (Thermo Fisher; 11875093). HCT116 and HT29 were cultured in McCoy's 5A Modified Medium (Thermo Fisher; 16600082). LoVo was cultured in Ham's F-12 (Thermo Fisher; 11765054), while LS174T was cultured in Minimum Essential Medium (Thermo Fisher; 11095080). All growth media were fortified to 10% FBS (Thermo Fisher; 16140071) and 2mM L-Glutamine (Thermo Fisher; 25030081). Cell lines were primarily grown in T75 Culture Flasks (Corning; 430641U) at 37°C in a humidified environment containing 5% CO_2. Cells were split upon reaching 70% confluence, approximately every two days. Present culture medium was aspirated, and cells were washed in 10mL of 1X PBS (Thermo Fisher; 10010023). Cells were dissociated from the cell culture flask by the addition of 3mL Trypsin-EDTA (Thermo Fisher; 25200056). During trypsinization, cells were incubated at 37°C in a humidified environment containing 5% CO_2. Following trypsinization, 5mL of fresh culture medium was added and cells were centrifuged at 1000 rpm for 5 minutes. The supernatant was aspirated and the cell pellet was resuspended in 5mL of fresh culture medium. Approximately 1mL of the cell suspension was added to a fresh T75 flask containing 9mL of the appropriate culture medium. Cell line identity, assessed via STR Profiling, and the absence of mycoplasma contamination, assessed via PCR (Genlantis; MY01050), was confirmed every six months.

3.1.2 Inverted siRNA Transfections

Cells were seeded in a 6-well tissue culture plate (Corning; 353046) at a density of 150,000 cells per well. The wells contained 2mL of RPMI-1640 growth medium fortified to 10% FBS and 2mM L-Glutamine. Silencer Select siRNAs (Thermo Fisher), Negative Control #1 (4390843) and siTCF7L2-s13880 (4392420), were delivered to the cells using the Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher; 13778150) and Opti-MEM Medium (Thermo Fisher; 31985070), according to the manufacturer's instructions. This resulted in the addition of 7.5µL of RNAiMAX and 25pmol of siRNA per well of the 6-well plate. The time series was constructed thus: 0- and 72-hour time points were transfected with siNEG and siTCF7L2, respectively, 24 hours after seeding. The 48-hour time point was transfected, with siTCF7L2, 24 hours after the 0- and 72-hour time points. The 24-hour time point was transfected, with siTCF7L2, 24 hours after the 48-hour time point. The cells were harvested 24 hours later, resulting in exposure to siTCF7L2 for 72, 48, and 24 hours. The 0-hour time point spent 0 hours in siTCF7L2 and serves as a transfection control for the siTCF7L2 72-hour time point. The siRNA:Lipofectamine complexes were not removed during the time course. This

process is illustrated in Figure 12, which demonstrates the advantages of the inverted transfection over the conventional transfection in terms of cell cycle homogeneity.

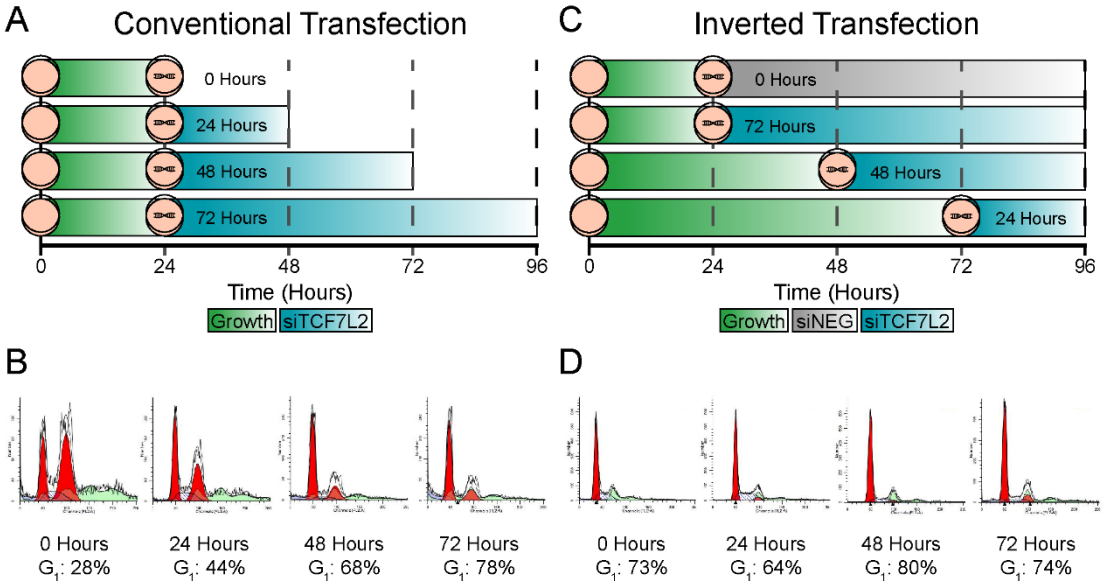

Figure 12 | Comparison of conventional and inverted transfection time series methodologies. In A, the conventional transfection methodology is shown, which consists of the simultaneous addition of siRNA to multiple cell culture vessels. The time course structure is generated by successive harvesting of the cells at defined time points from different vessels. Cells harvested at different time points are grown for different amounts of time, resulting in varying cell cycle distributions for each time point, shown in B. The G_1 peak is the leftmost peak in red, the G_2 peak is the rightmost peak in red, which decreases in size as time progresses. S-phase is denoted by the angled lines against a clear background found between the G_1 and G_2 peaks. Cellular debris is shown in purple and aggregates in light green. Given that the number of cells in G_1 varies from ~20% at Time 0 to ~80% at Time 72, we find this transfection methodology unsuitable for Hi-C analysis as differences in chromatin structure between the time points arise as a result of varying cell cycle distributions. Subfigure C shows the Inverse Transfection methodology, in which the cells are grown equal amounts of time, with successive addition of siRNA at defined time points. Cells are then harvested simultaneously. The time course was performed until 72 hours post-transfection based on the silencing efficiency, the capability to synchronize cells by this time point, as well as the results of a preliminary gene expression microarray, which demonstrated that the majority of changes in gene expression found at a later time point (96 hours post-transfection) also appeared at 72 hours post-transfection (results not shown). The inverted transfection methodology resulted in ~70% of the cells in the G_1/G_0 phase of the cell cycle across the time points, seen in D, thereby permitting time series Hi-C analysis. The population of cells in G_1 is shown in red and is the leftmost peak on the graph.

3.1.3 Dual-Luciferase Assay

SW480 cells were seeded in an opaque, white 96-well tissue culture plate (Corning; 3917) at a density of 5,000 cells per well. Cells were transfected 24 hours later with the Cignal TCF/LEF Reporter Assay Kit (Qiagen; CCS018L), using Lipofectamine 3000 (Thermo Fisher; L3000008) according to the manufacturer's instructions. Cells were simultaneously transfected with siTCF7L2 or siNEG, according to the Inverted Transfection protocol, using RNAiMAX. Cell lysates were prepared after the time course had been completed using the Dual-Luciferase Reporter Assay System (Promega; E1910), based on the manufacturer's instructions for Passive Lysis. Twice the recommended amount of 1X PLB (40µL) was added per well. Luciferase luminescence was measured according to the Dual-Luciferase Reporter Assay System instructions, using a Centro XS^3 LB 960 Microplate Luminometer (Berthold Technologies). Specifically, 100µL of LAR II was injected into a well, followed by a 2 second delay, followed by a 10 second measurement period. Immediately thereafter, 100µL of Stop & Glo reagent was injected, followed by a 2 second delay, and concluded with a 10 second measurement period. Each biological replicate consisted of three wells containing TCF/LEF Reporter Plasmid, one well containing Negative Control, one well containing Positive Control, and one well containing Non-Transfected Control for each sample. Three biological replicates were performed. Fold change was calculated according to a technical note from Promega, using the median value of the three TCF/LEF Reporter Plasmid transfected wells.

3.1.4 Growth Assay

SW480 cells transfected with the EV, APC, APC 60, APC 70, or APC 80 plasmid were selected with neomycin and then seeded in a 6-well tissue culture plate (Corning; 353046) at a density of 20,000 cells per well. The wells contained 2mL of RPMI-1640 growth medium fortified to 10% FBS and 2mM L-Glutamine. Six wells per sample were plated. Every 24 hours, cells from one well were trypsinized, centrifuged, and counted, using a hemocytometer, to determine the number of cells per well. Trypsinized wells were inspected using a light microscope to ensure that all cells had successfully detached. The growth assay occurred for 144 hours or 6 days. The growth rate was determined by dividing the number of cells on any given day by the number of cells on the first day.

3.2 Molecular Biology Techniques

3.2.1 Quantitative Polymerase Chain Reaction (qPCR)

RNA was extracted from at least three independent inverted transfections using the RNeasy Mini Kit and on-column DNase treatment (Qiagen; 74104, 79254), according to the manufacturer's instructions. RNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher). Synthesis of cDNA was performed with 800ng of RNA using the Verso cDNA Synthesis Kit (Thermo Fisher; AB1453B), based on the manufacturer's instructions. The optional RT Enhancer was used and equal amounts of both the Anchored Oligo dT and Random Hexamers were added. The cycling protocol consisted of a 30 minute incubation at 42°C, a 2 minute incubation at 95°C, and concluded with a 4°C hold. Quantitative PCR was performed for *TCF7L2*, *LEF1*, *CEACAM1*, *CtBP1*, *CtBP2*, and *YWHAZ* on an ABI PRISM 7000 Sequence Detection System (Applied Biosystems) or on a Lightcycler 480 II (Roche) using *Power* SYBR Green PCR Master Mix (Thermo Fisher; 4367659), according to the manufacturer's instructions. The thermocycling protocol consisted of a 2 minute incubation at 50°C, which proceeded into a 10 minute incubation at 95°C, followed by 40, two-step cycles consisting of 15 seconds at 95°C and 1 minute at 60°C. The cycling protocol was concluded with a melting curve analysis. Fold change was determined using *YWHAZ* as the reference gene and the $2^{-\text{deltadeltaCT}}$ method. Primer sequences can be found in Table 3.

Primer	Sequence	T_m (°C)	Product (bp)
TCF7L2_qPCR_F	CCTCACGCCTCTTATCACGTA	57	145
TCF7L2_qPCR_R	AGGCGATAGTGGGTAATACGG	57	
LEF1_qPCR_F	CGATGACGGAAAGCATCCAG	57	336
LEF1_qPCR_R	CCACCCGGAGACAAGGGATA	59	
YWHAZ_qPCR_F	ACTTTTGGTACATTGTGGCTTCAA	57	94
YWHAZ_qPCR_R	CCGCCAGGACAAACCAGTAT	58	
CAECAM1_qPCR_F	GACTCAGGACACAACCTACCTG	58	208
CAECAM1_qPCR_R	GGTGTCCGGGCCATAGGTG	61	
CtBP1_qPCR_F	GTGCCACATCCTGAACCTGTA	58	218
CtBP1_qPCR_R	AAGGGTCGTAGAAGAGCACG	57	
CtBP2_qPCR_F	CAGCGGACTCTACCATCTGC	58	206
CtBP2_qPCR_R	AAAGGCCTTGGCTCGAACTG	59	

Table 3 | qPCR primer sequences, annealing temperatures (T_m), and product sizes used in this dissertation are given in the table.

3.2.2 Western Blot

Nuclear and cytoplasmic protein fractions were extracted using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher; 78833), according to the manufacturer's instructions. Protein was prepared for quantification using the Pierce BCA Protein Assay Kit (Thermo Fisher; 23225), according to the manufacturer's instructions, and quantified on a SpectraMax M2^{e} microplate reader (Molecular Devices). Western blot samples were generated with 30µg protein, NuPAGE LDS Sample Buffer (Thermo Fisher; NP0007), and Molecular Biology Grade Water (Corning; 46000CM). Samples were heated to 70°C for 10 minutes. The samples, 25µL in volume, were electrophoresed for 50 minutes at 200V through a 4-12% Bis-Tris gel (Thermo Fisher; NP0321BOX) in an XCell *SureLock* Mini-Cell (Thermo Fisher; EI0001) with MOPS Running Buffer (Thermo Fisher; NP0001). Proteins were transferred into an Immobilon-P PVDF Membrane (Millipore; IPVH00010) for 90 minutes at 30V using Transfer Buffer (Thermo Fisher; NP00061) and an XCell II Blot Module (Thermo Fisher; EI9051). Membranes were washed once for 5 minutes in TBS (Takara; T903) and blocked in TBS with 5% milk (VWR; M20310G). Membranes were then washed 3 times for 5 minutes each, shaking at room temperature, in TBS. Membranes were incubated overnight at 4°C in TBS with 5% BSA (Roche; 03117332001) and the primary antibody. Primary antibody dilutions recommended by the manufacturer were used (usually 1:1000). Then, membranes were washed 3 times for 5 minutes each in TBS, shaking, at room temperature. Membranes were then incubated with an HRP-linked secondary antibody in TBS for 1 hour at room temperature, shaking. Secondary antibody dilutions recommended by the manufacturer were used (1:2000). The membrane was then washed 3 times for 5 minutes each in TBS, shaking, at room temperature. Antibodies were detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher; 34080) for 5 minutes in the dark. Blots were imaged using an Azure c600 Gel Imaging Station (Azure Biosystems). Primary antibodies used in this study targeted TCF4 (CST; 2569S), LEF1 (CST; 2230S), β-catenin (CST; 9587S), Active β-catenin (Millipore; 05665), CtBP1 (CST; 8684S), CtBP2 (CST; 13256S), or TBP (CST; 8515S). The secondary antibodies used in this study were Goat Anti-rabbit IgG, HRP-linked and Horse Anti-mouse IgG, HRP-Linked (CST; 7074S, 7076S).

3.2.3 RNA Sequencing

RNA was extracted from three independent inverted transfections using the RNeasy Mini Kit and on-column DNase treatment (Qiagen; 74104, 79254), according to the manufacturer's instructions. RNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher). Silencing of *TCF7L2* or *LEF1* was determined using qPCR. RNA Integrity was determined by the Center for Cancer Research Genomics Core using the 4200 TapeStation (Agilent). RNA samples with an RNA Integrity Number (RIN) of 9 or higher were used for library preparation. Libraries were prepared and sequenced by the Center for Cancer Research Sequencing Facility. Libraries were prepared using the TruSeq Stranded Total RNA Library Prep Gold Kit (Illumina; 20020598), according to the manufacturer's instructions. Samples were sequenced on a HiSeq 3000 generating 125 base pair, paired-end reads.

Replicate concordance was determined using PCA. Replicates from the same time points clustered together and moved progressively over time.

3.2.4 Chromatin Immunoprecipitation (ChIP-seq)

Chromatin Immunoprecipitation followed by high-throughput sequencing, or ChIP-seq, was performed using the SimpleChIP Plus Enzymatic Chromatin IP Kit (CST; 9005S), based on the manufacturer's instructions. Reagent numbers, e.g., #7005, refer to a specific component of the SimpleChIP Kit.

DNA-protein crosslinks were formed by incubating SW480 cells for precisely 10 minutes in a 1% formaldehyde low-serum solution generated by adding 54µL of a 37% formaldehyde solution (Sigma; F8775-25ML) per well of a 6-well plate containing 2mL Opti-MEM Serum Reduced Medium (Thermo Fisher; 31985070). Immediately thereafter, 200µL of a 10X glycine solution (#7005) was added and the 6-well plate was swirled to mix the solutions and incubated for 5 minutes at room temperature. The solution was aspirated and cells were washed twice with 2mL of ice-cold 1X PBS (Thermo Fisher; 10010023). Then, 2mL of ice-cold PBS containing protease inhibitors (#7012) was subsequently added and the cells were scraped and collected into a 15mL conical tube. Cells were centrifuged at 1,500 rpm for 5 minutes at 4°C and the resulting supernatant was aspirated. Cells were resuspended in 3mL of ice-cold Buffer A (#7006) supplemented with DTT (#7016) and protease inhibitors, then incubated on ice for 10 minutes with inversion every 3 minutes. The suspension was centrifuged at 2,500 rpm for 5 minutes at 4°C and the supernatant was aspirated. The pellet was then resuspended in 3mL of ice-cold Buffer B (#7007) supplemented with DTT. The centrifugation was repeated, the supernatant was aspirated, and the pellet was resuspended in 300µL of ice-cold Buffer B, supplemented with DTT. Then, 1µL Micrococcal Nuclease (#10011) was added to the 300µL of nuclei and the mixture was incubated at 37°C for 20 minutes, with mixing by flicking every 4 minutes. The digestion was stopped by the addition of 30µL of 0.5M EDTA (#7011) and a 2 minute incubation on ice. The nuclei were pelleted by centrifugation at 16,000 rcf for 1 minute at 4°C, and the supernatant was removed. The pellet was resuspended in 300µL 1X ChIP Buffer (#7008) containing protease inhibitors and incubated on ice for 10 minutes. The lysate was then sonicated with a Branson 450 Sonicator (Branson Ultrasonics Corporation) for 10 seconds at 10% capacity. The use of a Dounce homogenizer for nuclei homogenization was entertained and found to be useless, we recommend against its use. The lysates were then centrifuged at 9,400 rcf for 10 minutes at 4°C. The supernatant was then transferred to a DNA low-binding tube (Eppendorf; 022431021) and 30µL were removed to analyze chromatin digestion and yield.

To determine chromatin yield and the extent of digestion, 120µL of Molecular Biology Grade Water (Corning; 46000CM), 6µL of 5M NaCl (#7010), and 2µL of RNase A (#7013) were added to each of the 30µL chromatin samples. The samples were mixed well by pipetting up-and-down and incubated at 37°C for 30 minutes. Following incubation, 2µL of Proteinase K (#10012) was added, the samples were mixed well, and the reaction was incubated at 65°C for 2 hours. The DNA was then purified using the supplied DNA Purification Columns and the purification procedure described at the

end of this methodology. The purified DNA was run on a 1.5% agarose gel and the concentration was tested using a NanoDrop 1000 spectrophotometer (Thermo Fisher). The average concentration per sample was approximately 150ng/μL. The chromatin was digested to fragments ~150, 300, 450, 600, 750, and 900bp in length. The majority of the chromatin was either 150 or 300bp in length.

For chromatin immunoprecipitation, 7.5μg of digested, cross-linked chromatin was diluted in sufficient 1X ChIP Buffer and protease inhibitors to bring the total reaction volume to 500μL. A 10μL aliquot was removed, which represents the 2% Input Sample. Antibodies targeting TCF4 (Millipore; 1710109) and LEF1 (Millipore; 17604). The amount of each antibody per reaction was: TCF4 – 5μg and LEF1 – 4μg. The samples were then incubated overnight at 4°C with rotation, using a Mini Lab Roller Dual Format Rotator (Labnet).

ChIP-Grade Protein G Magnetic Beads (#9006) were resuspended by gentle vortexing and pipetting up-and-down. Immediately thereafter, 30μL of beads were added to each IP reaction and incubated at 4°C for 4 hours with rotation. The magnetic beads were then separated by placing the tubes in a magnetic separation rack (CST; 14654) for 2 minutes. The supernatant was then carefully removed, and the magnetic beads were resuspended in 1mL of low salt wash and incubated for 5 minutes at 4°C with rotation. Beads where then separated by placing tubes in the magnetic rack and the supernatant was carefully removed. The low salt wash procedure was repeated two more times, for a total of three low salt washes. After the final low salt wash, the beads were resuspended in 1mL of high salt wash and incubated for 5 minutes at 4°C with rotation. Beads were then separated in the magnetic separator and the supernatant was carefully removed.

To elute chromatin from the antibody/protein G magnetic beads complex, 150μL of 1X ChIP Elution Buffer (#7009) was added to each IP sample. The IP samples were then incubated at 65°C for 30 minutes with gentle vortexing (1200 rpm). The magnetic beads were collected by placing the tubes in the magnetic separator and the supernatant was carefully transferred to a new DNA low-binding tube. For crosslink reversal, 150μL of 1X ChIP Elution Buffer was added to the 2% Input Sample. Crosslinks were reversed by adding 6μL 5M NaCl and 2μL Proteinase K to all tubes (including the Input samples), and incubating overnight at 65°C.

The DNA was then purified by adding 750μL of DNA Binding Buffer (#10007) to each DNA sample and pipetting up-and-down. Then, 450μL of each sample was transferred to a DNA Purification Column (#10010). The columns were centrifuged at 14,000 rcf for 30 seconds to bind the DNA to the column. The flowthrough was discarded and the remaining 450μL of each sample was transferred to the DNA Purification Column. The centrifugation step was repeated and the flowthrough was discarded. Then, 750μL DNA Wash Buffer (#10008) was added to the DNA Purification Column and centrifuged at 14,000 rcf for 30 seconds to wash the column. The flowthrough was discarded. The centrifugation step was repeated, at 14,000 rcf for 30 seconds, to collect all remaining liquid in the column. The column was transferred to a new collection tube and 50μL of DNA Elution Buffer (#10009) was added. The column was incubated at room temperature for 10 minutes. Elution was completed by

centrifugation at 14,000 rcf for 1 minute. The used DNA Purification Column was subsequently discarded.

Immunoprecipitated DNA quality and size distribution was determined using an Agilent TapeStation 4200, performed by the Center for Cancer Research Genomics Core. Samples which passed the quality control threshold were sent to the Center for Cancer Research Sequencing Facility for library preparation and sequencing. Libraries were prepared using the TruSeq ChIP Library Preparation Kit (Illumina; IP2021012), according to the manufacturer's instructions. Samples were sequenced on a HiSeq 4000 generating 125 base pair, paired-end reads.

3.2.5 Chromosome Conformation Capture (Hi-C)

The *in situ* Hi-C protocols from Rao et al. were adapted with slight modifications[177]. SW480 cells, which had undergone the inverse transfection time series, were crosslinked in a 1% formaldehyde solution (Sigma; F8775-25ML) for 10 minutes. Formadehyde was quenched by the addition of glycine (Sigma; 50046-250G). For each Hi-C library, approximately five million cells were incubated in 250µL of ice-cold Hi-C Lysis Buffer (Corning; 46-000-CM, Thermo Fisher; 15568025, Thermo Fisher; AM9760G) with 50µL of protease inhibitor cocktail (Sigma; P8340-1ML) on ice for 30 minutes and washed with 500µL of Hi-C Lysis Buffer. The nuclei were pelleted by centrifugation at 2500 rcf for five minutes at 4°C, resuspended in 50µL of 0.5% sodium dodecyl sulfate (SDS)(Sigma; 71736-100ML) and incubated at 62°C for 10 minutes. Afterwards, 145µL of water and 25µL of 10% Triton X-100 (Sigma; 93443-100ML) were added and incubated at 37°C for 15 minutes.

Chromatin was digested with 100 units of MboI (NEB; R0147S) overnight at 37°C with rotation. Chromatin end overhangs were filled in and marked with biotin-14-dATP by adding the following reagents to the reaction: 37.5µL of 0.4mM biotin-14-dATP (Thermo Fisher; 19524016), 1.5µL of 10mM dCTP (NEB; N0446S), 1.5µL of 10mM dGTP (NEB; N0446S), 1.5µL of 10mM dTTP (NEB; N0446S), and 8µL of 5U/µL DNA Polymerase I, Large (Klenow) Fragment (NEB; M0210S). The marked chromatin ends were ligated by adding 900µL of ligation master mix consisting of 663µL of water, 120µL of 10X NEB T4 DNA Ligase Buffer (NEB; B0202S), 100µL of 10% Triton X-100, 12µL of 10mg/mL BSA (Roche; 03117332001) and 5µL of 400U/µL T4 DNA Ligase (NEB; M0202S) and were then incubated at room temperature for four hours.

Crosslinking was reversed by adding 50µL of 20mg/mL proteinase K (NEB; P8107S) and 120µL of 10% SDS followed by an incubation at 55°C for 30 minutes. Then, 130µL of 5M sodium chloride was added and crosslinks were reversed at 68°C overnight. DNA was precipitated with ethanol, washed with 70% ethanol (Sigma; E7023-1L), and dissolved in 130µL of 10mM Tris-HCl, pH 8.0. DNA was sheared on a Covaris S2 sonicator. Biotinylated DNA fragments were pulled down with MyOne Streptavidin T1 beads (Thermo Fisher; 65601). To repair the ends of sheared DNA and remove biotin from unligated ends, DNA-bound beads were resuspended in 100µL of mix containing 82µL of 10X NEB T4 DNA Ligase Buffer, 10µL of 25mM dNTP Mix, 5µL of 10U/µL NEB T4 PNK (NEB; M0201S), 4µL of 3U/µL T4 DNA Polymerase (NEB; M0203S), and 1µL of 5U/µL DNA polymerase I, large (Klenow) fragment. After end-repair, dATP

attachment was carried out in 100µL of mix consisting of 90µL of 1X NEBuffer 2 (NEB; B7002S), 5µL of 10mM dATP, 5µL of 5U/µL Klenow exo- minus (NEB; M0212S), and incubated at 37°C for 30 minutes. The beads were then cleaned for Illumina sequencing adaptor ligation which was done in a mix containing 50µL of 1X NEB Quick Ligation Reaction Buffer (NEB; B6058), 3µL of NEB DNA Quick Ligase (NEB; M2200S), and 2µL of a 15µM Illumina indexed adapter (Illumina; 20015960) at room temperature for 1 hour. DNA was dissociated from the beads by heating at 98°C for 10 minutes, separated on a magnet, and transferred to a clean tube. Final amplification of the library was carried out in multiple PCR reactions using Illumina PCR primers. The reactions were performed on a 25µL scale consisting of 25ng of DNA, 2µL of 2.5mM dNTPs, 0.35µL of each primer at a 10µM concentration, 2.5µL of 10X PfuUltra Buffer (Agilent; 600672), and 0.5µL of PfuUltra II Fusion DNA polymerase (Agilent; 600672). The PCR cycle conditions were set to 98°C for 30 seconds as the denaturing step, followed by 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30 seconds with a final extension step at 72°C for 7 minutes. After PCR amplification, products from the same library were pooled and fragments ranging from 300-500 bp were selected using AMPure XP beads (Beckman Coulter; A63880). The size-selected libraries were sequenced on a HiSeq 2500 using Illumina TruSeq v4 chemistry generating 125 base pair, paired-end reads.

3.3 Bioinformatics

3.3.1 Data Retrieval – GTEx

GTEx data was downloaded from the GTEx Dataset Portal (https://gtexportal.org/home/datasets). The most recent RNA Sequencing data release, GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.0_gene_tpm.gct.gz, which was aligned to reference genome hg38, was used.

3.3.2 Data Retrieval – TCGA

TCGA data was retrieved using the TCGAbiolinks (Version 2.16.4) package in R. The newest release, which was aligned to reference genome hg38, was used.

3.3.3 RNA Sequencing Data Analysis

RNA sequencing data was processed using the CCBR Pipeliner Version 4.0.2. Briefly, FASTQC was used to assess sequencing quality and Cutadapt was used to remove adapter sequences and perform quality trimming, respectively. Kraken, KronaTools, and FastQScreen were used to assess microbial contamination. Our samples were confirmed to be free of microbial contamination. STAR (two-pass) was used to align reads to the hg38 reference genome. Picard, Preseq, SAMtools, and RSeQC were used to assess alignment quality. Gene expression was quantified using RSEM. Transcripts per million (TPM) were computed to normalize gene expression values. Differential gene expression was performed using limma. Volcano plots were generated using the EnhancedVolcano R package.

3.3.4 Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis was performed using R and the fgsea package, available through Bioconductor. Differentially expressed genes, as calculated using limma, were pre-ranked according to their $\log_2$ fold change values. Gene set enrichment was calculated using the fgsea command and the Hallmark gene sets available from MSigDB. The minSize and maxSize parameters were set to 15 and 500, respectively. The heatmap was generated using the Normalized Enrichment Scores, generated by fgsea, and the ComplexHeatmap R package[178].

3.3.5 ChIP Sequencing Data Analysis

ChIP sequencing data was processed using the CCBR Pipeliner Version 4.0.2 and aligned to the hg38 reference genome. The pipeline employs FASTQC, FastQScreen, Kraken, and KronaTools to assess sequencing quality and microbial contamination. Our samples were confirmed to be free of microbial contamination. DeepTools was used to generate a fingerprint plot and Picard was used to asses duplicates (~90% read pairs were unique). Samples passed ChIP-seq specific metrics such as NRF, NSC, PBC1, PBC2, and RSC and contained at least 32 million uniquely mapped reads

per sample. Peaks were identified using MACS2 narrow. Input-normalized bigwig files were loaded into IGV (version 2.8.2) for visualization.

3.3.6 Hi-C Data Analysis

Hi-C data analysis was performed in its entirety by Gabrielle Dotson from the Laboratory of Dr. Indika Rajapakse at the University of Michigan, Ann Arbor.

Chapter 4: TCF4 and LEF1 WNT Transcription Factors in Colon Cancer

4.1 Interactions of the WNT Signaling Transcription Factors

4.1.1 WNT Transcription Factor Expression in Non-Diseased Colon

Canonical WNT signaling activity is crucial for colon homeostasis as well as for the initiation and maintenance of colon cancer. When WNT signaling is active, β-catenin migrates to the nucleus where it binds the TCF/LEF family of transcription factors, which transmits the canonical WNT signal. Since each of these factors can interact with β-catenin, yet each initiates a different transcriptional program, it is important to understand which TCF/LEF factors are expressed in normal and cancerous colon tissue. Understanding the expression patterns of these factors establishes a foundation for understanding WNT signaling output in colon tissue.

To understand the expression levels of the four WNT signaling transcription factors in the non-diseased human colon, I downloaded data from the open-access Genotype-Tissue Expression (GTEx) project. GTEx is an ongoing effort designed to generate a comprehensive resource of tissue-specific gene expression and regulation data collected from nearly 1,000 individuals across 54 non-diseased tissues, including the colon. Data was downloaded from the GTEx website, loaded into R, and samples from the sigmoid and transverse colon were selected. Several steps were taken to ensure proper data quality: patient samples with severe autolysis were removed, samples with an exonic rate less than 80% were removed, and samples must have been sequenced using the TruSeq.v1 sequencing chemistry. In some instances, multiple samples corresponded to a single patient, as samples were extracted at multiple sites.

Given the intention of identifying breadth of expression of the WNT transcription factors, only one sample per patient was selected. This resulted in a total of 379 samples split evenly between the sigmoid and transverse regions (47% and 53%).

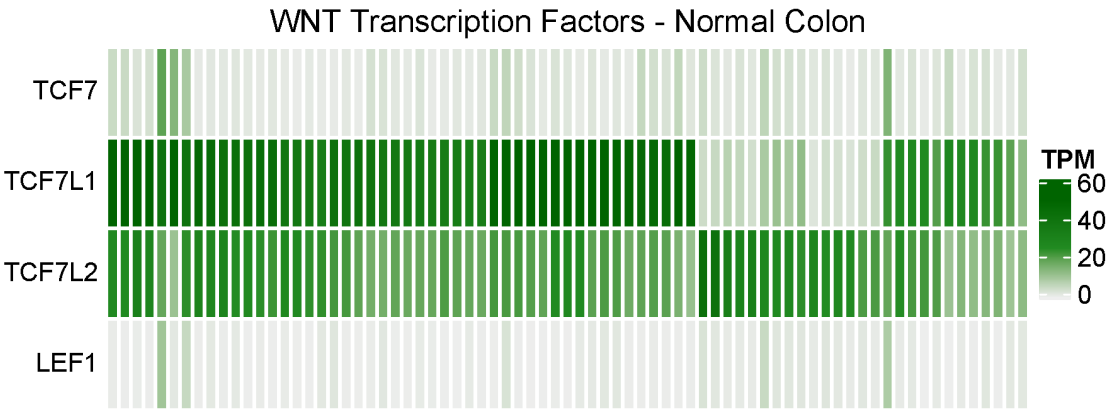

Figure 13 | WNT TF Expression, plotted as rows, from GTEx with samples plotted as columns (n=75).

The expression of the four WNT signaling transcription factors in a representative subset consisting of 75 patient samples was then plotted (Figure 13). The subset was plotted to facilitate inspection of the expression, plotted as transcripts per million (TPM), of the WNT transcription factors within a sample. As has been previously reported in the literature, *TCF7L2* is the dominant WNT signaling transcription factor in the colon and is expressed in 100% of samples (379/379). *TCF7L1* is expressed in 87% of the samples (329/379). *TCF7* and *LEF1* are not appreciably expressed across the normal colon samples, at 8% and 1%, respectively. In terms of expression intensity, as opposed to breadth, *TCF7* and *LEF1* are lowly expressed at ~3 and 1.5 TPM, while *TCF7L2* and *TCF7L1* are strongly expressed, at ~21 and 30 TPM, respectively (Figure 14).

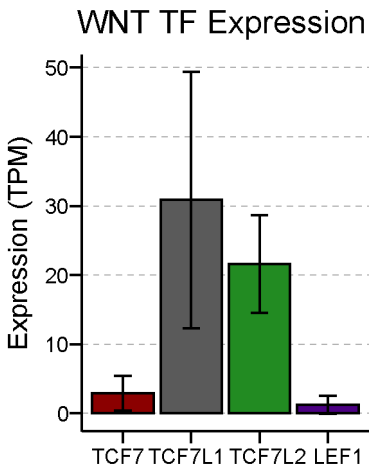

Figure 14 | Average WNT TF Expression from GTEx (n=379). Error bars show the standard deviation.

4.1.2 WNT Transcription Factor Expression in Colon Cancer

To determine the expression of the four WNT signaling transcription factors in colon cancer, I downloaded data from The Cancer Genome Atlas (TCGA). TCGA is a landmark cancer genomics program which molecularly characterized over 20,000 tumors across 33 cancer types. The data is publicly available and was accessed using the R package, TCGAbiolinks (2.16.4). RNA sequencing data was queried from the "TCGA-COAD" project. Samples generated with the same workflow method, HTseq, and normalized to fragments per kilobase million (FPKM) were selected. Instances in which multiple samples corresponded to a single patient, were, as previously, removed to limit one sample per patient. This resulted in a total of 254 samples, which were derived from four regions: the ascending (26%), transverse (12%), descending (6%), and sigmoid colon (56%). While not an even split, the sample distribution follows the distribution of colon cancer found in humans, with higher tumor prevalence in the left colon (descending and sigmoid), than in the right colon (ascending and transverse).

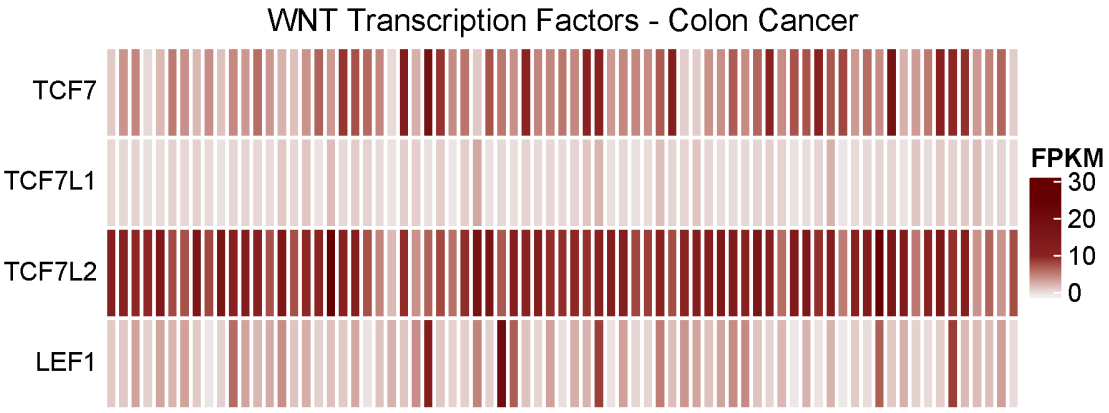

Figure 15 | WNT TF Expression, plotted as rows, from TCGA with samples plotted as columns (n=75).

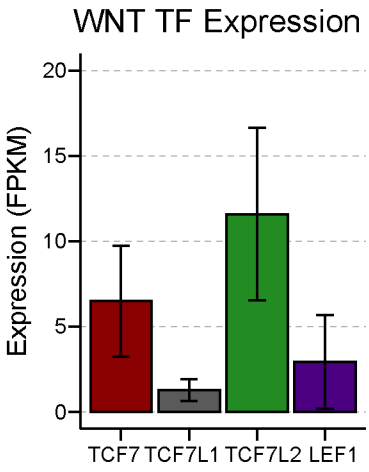

Figure 16 | Average WNT TF Expression from TCGA (n=254). Error bars show the standard deviation.

The expression of the four WNT signaling transcription factors in a representative subset consisting of 75 patient samples was plotted (Figure 15). *TCF7L2* remains the dominant transcription factor in colon cancer and is expressed in 100% of the samples (254/254), while the expression of *TCF7L1* is observed in only 3% of the samples (8/254). The expression of *TCF7* and *LEF1* is evident in 90% and 42% of samples, respectively. Across the samples, *TCF7L2* remains the highest expressed WNT transcription factor at ~12 FPKM, while expression of *TCF7L1* is the lowest at ~1.3 FPKM. The expression of *TCF7* and *LEF1* is ~7 and 3.5 FPKM, respectively, which is approximately one-half the expression of *TCF7L2* (Figure 16). Average transcription factor expression was calculated from all 254 TCGA samples.

A confounding factor in the TCGA data is tumor purity, as some cancers may have large numbers of tumor-infiltrating lymphocytes. The inclusion of these cells in samples for RNA sequencing may result in artificially increased levels of immune system centric genes, such as *TCF7* and *LEF1*. To determine the influence of tumor purity on WNT transcription factor expression, I obtained various measures of tumor purity from TCGAbiolinks including ESTIMATE, ABSOLUTE, LUMP, IHC, and CPE. ESTIMATE is based on gene expression profiles of 141 immune genes and 141 stromal genes, ABSOLUTE uses somatic copy-number data, LUMP is based on methylation data from 44 non-methylation immune-specific CpG sites, IHC is an estimate based on image analysis of haematoxylin and eosin stained slides, and CPE

is a consensus measurement of the median purity level after normalizing levels from all methods to give them equal means and standard deviation.

Purity Metric	TCF7	LEF1
ESTIMATE	0.09	-0.15
ABSOLUTE	0.15	0.01
LUMP	0.12	-0.08
IHC	-0.01	-0.07
CPE	0.28	-0.08

Table 4 | Pearson correlation between *TCF7* or *LEF1* expression and tumor purity metrics.

I found no correlation between the expression of *TCF7* or *LEF1* with any of the purity measures, confirming that the expression of *TCF7* and *LEF1* can not be attributed to the presence of tumor-infiltrating lymphocytes. Interestingly, ESTIMATE had the strongest correlation with *LEF1* (Table 4). ESTIMATE is based on the expression of immune and stromal genes. If a cancer ectopically expresses an immune system regulating transcription factor, the expression of immune system genes would likely increase, regardless of invading lymphocytes. Tumor purity, as determined from ESTIMATE, should therefore be interpreted with caution and compared with other measures of tumor purity (such as ABSOLUTE and IHC).

4.1.3 WNT Transcription Factor Expression in Transverse and Sigmoid Colon

The most dramatic difference in WNT transcription factor expression occurred in *TCF7L1*, whose expression breadth was 87% of non-diseased samples yet only 3% of samples in colon cancer. In colon cancer samples where *TCF7L1* is present, it is expressed at a low degree. *TCF7L1* is typically recognized as a repressive WNT factor and its loss may be required for colon cancer cells to take advantage of activated WNT signaling. Given the potential connection between *TCF7L1* expression and colon cancer, I checked if the expression of the WNT transcription factors varied between

regions within the non-diseased colon (transverse and sigmoid), since colon cancer has been consistently reported to occur more frequently in the left colon. This would provide insight into whether an expression bias between different regions of the normal colon (left and right) may contribute to colon cancer incidence (Figure 17).

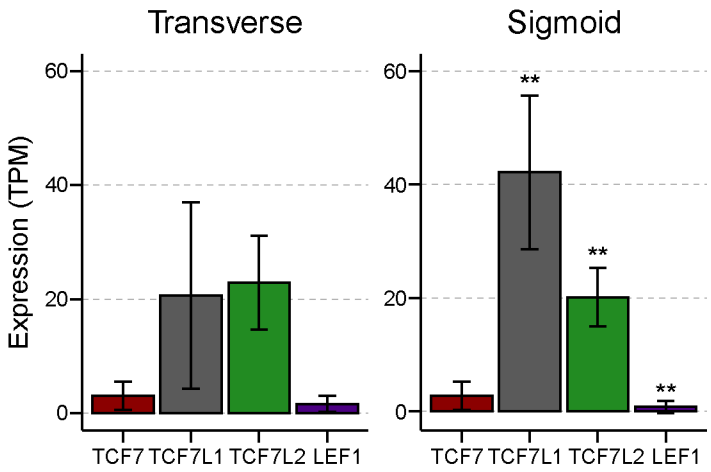

Figure 17 | WNT TF expression in the Transverse and Sigmoid Colon from GTEx (n=199, 180). The bars show the average and the error bars denote the standard deviation. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

The expression of *TCF7L1*, *TCF7L2*, and *LEF1* are all significantly different in the sigmoid and transverse colon when using Student's two-sided t-test ($p < 0.01$). This is driven by the large number of samples. The difference in expression of 2 TPM observed in *TCF7L2* and *LEF1* likely has no impact on biological function. The higher expression levels of *TCF7L1* in the sigmoid colon, may influence colon cell behavior and may play a protective role against colon tumorigenesis in the sigmoid colon.

4.1.4 WNT Transcription Factor Expression Ratios

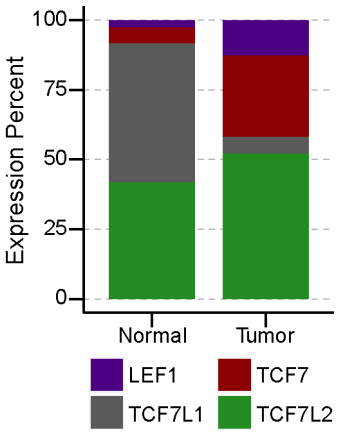

Figure 18 | WNT TF Percent Expression (n=379, 254).

Since all WNT transcription factors can interact with β-catenin, the canonical WNT activator, the relative expression of the four WNT transcription factors was determined to explore how WNT signaling output may shift as the ratio of the WNT signaling transcription factors changes during the progression from normal to cancer. The percent expression of each WNT transcription factor was calculated across all GTEx (379) and TCGA (254) samples and averaged to estimate the relative abundance of each WNT transcription factor (Figure 18). I found that in normal tissue, *TCF7L1* and *TCF7L2* are the two dominant transcription factors through which the WNT signal is most likely to be relayed. The expression of *LEF1* and *TCF7* accounted for ~8% of total WNT transcription factor expression in normal tissue. This arrangement shifts dramatically in colon cancer, where *LEF1* and *TCF7* now account for ~43% of WNT transcription factor expression, while *TCF7L2* accounts for 52%. *TCF7L1*, its expression almost entirely lost, now accounts for 6% of total WNT transcription factor expression. How does a different balance of WNT transcription factor expression influence downstream target gene expression? Given that colon cancers have activated WNT signaling activity, and that expression of *TCF7L2* is still high, the likelihood that LEF1 and TCF1 drive a similar transcriptional program to TCF4 is unlikely as further up-regulating already activated target genes carries little functional benefit. Instead, given that LEF1 and TCF1 are primarily involved in the activation of the expression of immune system genes in adults, which differs from the role of TCF4, one anticipates increased

expression of lymphoid genes. However, since all of the WNT transcription factors can bind the cognate WNT motif (5'-SCTTTGATS-3'), through their highly conserved HMG domain, there may be functional overlap between the target genes of LEF1, TCF1, and TCF4, resulting in synergistic expression of WNT target genes, in addition to the up-regulation of lymphatic genes by LEF1 and TCF1 alone. Additionally, *TCF7*, *LEF1*, and *TCF7L2* have been reported to be WNT target genes themselves, so a shift in WNT signaling activity could perturb their expression pattern. We therefore decided to study the impact of *TCF7L2*, the dominant WNT signaling transcription factor in the colon, on transcriptional dynamics and nuclear structure in colon cancer.

4.2 The Influence of *TCF7L2* on Transcriptional and Structural Dynamics

4.2.1 *TCF7L2* Silencing

TCF7L2 was silenced across three time points, each 24 hours apart, by successive transfection of a small interfering RNA (siRNA) into SW480 colon cancer cells. Initially, the efficiency of three different siRNAs, each binding a different region of *TCF7L2*, were tested. The most effective siRNA targeted the 3' UTR of *TCF7L2* and therefore, was capable of down-regulating nearly all variants of TCF4. This siRNA was used throughout the study. To minimize confounding factors based on varying cell cycle distributions, I developed an inverse transfection protocol, which resulted in ~70% of the cell population in G_1 at harvesting. Conventional transfection methodologies result in varying cell cycle distributions with ~20% of cells in G_1 at the first time point, while ~80% of cells are in G_1 at the last time point. A detailed description and diagram can be found in the Methods chapter (Figure 12).

Silencing of *TCF7L2* was confirmed at the RNA level using qPCR, which demonstrated a 4-fold reduction in *TCF7L2* transcript abundance by Time 72 (Figure 19). *TCF7L2* transcript abundance was normalized to the expression of the reference gene *YWHAZ*. Loss of TCF4 protein in the nuclear fraction of cell lysates was established using western blot, with a 5% decrease by Time 24, an 80% decrease by Time 48 and a 97% decrease by Time 72 (Figure 19). The western blot demonstrates down-regulation of multiple isoforms (78 kDa and 58kDa) of TCF4. The TATA Binding Protein (TBP) was used as loading control and TCF4 was not detected in the cytoplasmic fraction (results not shown).

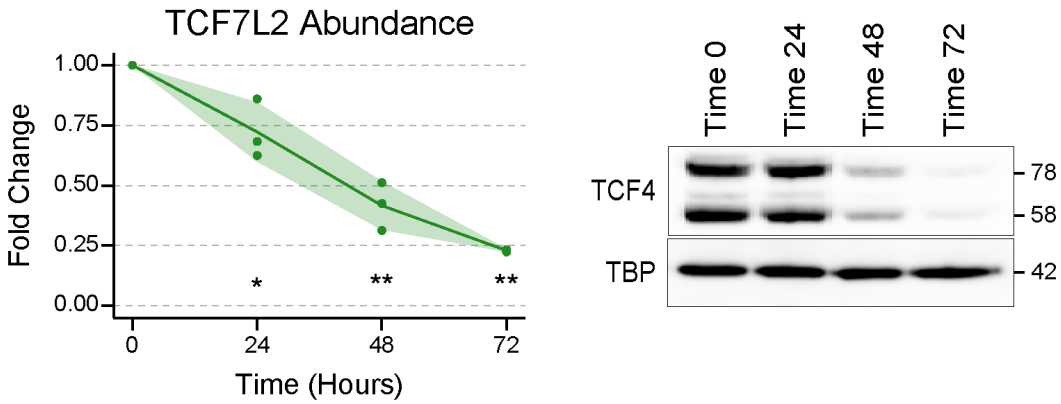

Figure 19 | *TCF7L2* transcript abundance was assayed with qPCR against the reference gene *YWHAZ* (left). Three biological replicates are plotted for each time point, shown as green dots (n=3). The green line represents the average and the shaded ribbon denotes the standard deviation. Statistical significance was calculated using Student's two-sided t-test (*p < 0.05, **p < 0.01). TCF4 abundance was determined via western blot (right) with TBP as loading control.

Following successful silencing of *TCF7L2*, RNA sequencing was performed. RNA was harvested from *TCF7L2*-silenced cells and treated with DNase I to degrade potential genomic DNA. RNA integrity was determined by the CCR Genomics Core using an Agilent TapeStation 4200. Samples with an RNA Integrity Number (RIN) of 9 or higher were sent to the CCR Sequencing Facility for library preparation and sequencing.

4.2.2 A Disproportionate Up-Regulation in Gene Expression

Silencing of *TCF7L2* resulted in a progressive change in the transcriptome over the time series. Given that TCF4 is required for maintaining colon cell homeostasis and that WNT signaling is constitutively activated in the SW480 colon cancer line, the anticipated role of TCF4 is one of a transcriptional activator alongside β-catenin. However, we observe a disproportionate up-regulation of gene expression following the silencing of *TCF7L2*, suggesting that TCF4 may function simultaneously as a transcriptional repressor. This is not entirely unexpected, given that TCF4 can interact

with transcriptional repressors, though this is thought not to occur in cells with high amounts of nuclear β-catenin. Differential gene expression was calculated using limma and visualized on volcano plots, which show the $\log_2$ fold change as well as the adjusted *p*-value (FDR) for each gene (Figure 20). At 48 hours post-transfection, 124 genes were significantly up-regulated and had a $\log_2$ fold change in expression of two or greater when compared to the control (siNeg treated cells). Four times the number of genes were up-regulated as compared to down-regulated at 48 hours and approximately twice the number of genes were up-regulated as down-regulated at 72 hours post-transfection. Results at 24 hours post-transfection were not interpreted due to the negligible reduction in TCF4 levels observed in the western blot.

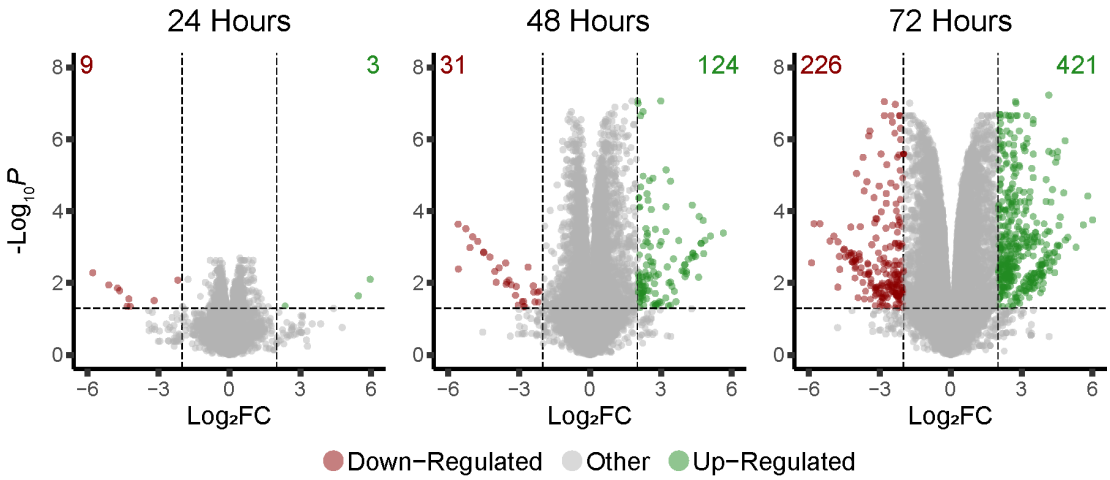

Figure 20 | Differential gene expression in *TCF7L2*-silenced SW480 cells. The threshold for $\log_2$ fold change was set to two and the threshold for the FDR adjusted *p*-value was set to 0.05. The genes which exceeded the fold change threshold and were below the *p*-value threshold are colored green for up-regulated and red for down-regulated. The number of genes in each category are shown in the corresponding color. Two biological replicates were used (n=2).

To explore the biological relevance of the transcriptomic changes, I used Gene Set Enrichment Analysis (GSEA). GSEA uses gene sets, comprised of genes associated with a biological function, and a list of genes ranked by expression to compare which gene sets are more active, based on the expression rank of the genes they contain. An enrichment score is then generated, and normalized based on the size of the pathway, yielding a normalized enrichment score (NES). I used the Hallmark Genesets, which are a set of 50 commonly used, publicly available gene sets available from the Molecular Signatures Database (MSigDb). The NES for all 50 pathways were calculated for each time point and the top ten up-regulated and the top ten down-regulated pathways, based on their adjusted *p*-value at the last time point (72 hours), were selected (Figure 21). Up-regulated pathways included immune response

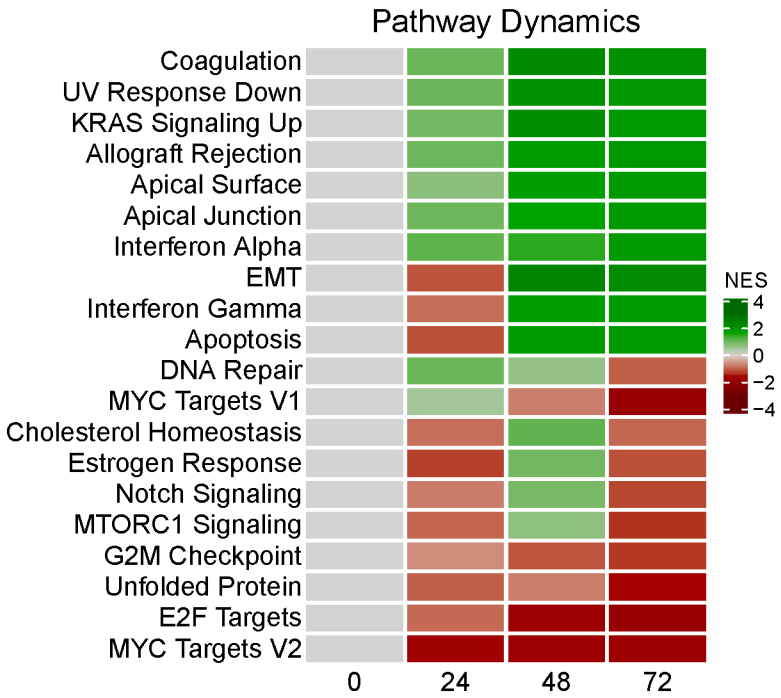

Figure 21 | GSEA for *TCF7L2*-silenced SW480 cells. The NESs for all 50 Hallmark gene sets were calculated and the top ten up-regulated as well as the top 10 down-regulated pathways, based on their adjusted *p*-value at the last time point, were plotted.

pathways such Interferon Alpha, Interferon Gamma, and Allograft Rejection, while down-regulated pathways included MYC Targets V2, MYC Targets V1, as well as E2F Targets. The down-regulation of the MYC gene sets is anticipated, since TCF4 is a potent driver of *MYC* expression in the colon. E2F is also logical as previous studies have reported that interfering with TCF4 function results in accumulation of cells in G_1 due to a block at the G_1/S transition. E2F targets are known to facilitate progression of cells into S phase, which may explain the G_1/S block. This may work in concert with MYC, as MYC also facilitates entry into S phase. The expression enrichment in most pathways is significant, with the exception of the DNA Repair, Cholesterol Homeostasis, Estrogen Response, Notch Signaling, MTORC1 Signaling, and G2M Checkpoint pathways., which did not display a significant change in activity.

Closer inspection of the changes in pathway activity shows that the Apoptosis, Interferon Gamma and EMT pathways respond in a similar fashion, with an initial decrease in activity at the 24-hour time point followed by a strong increase in enrichment at the 48- and 72-hour time points. Similarly, the MTORC1, Notch, Estrogen Response and Cholesterol pathways share a similar response to *TCF7L2* silencing. This coordinated response suggests that these pathways may function cooperatively when the cell is responding to a specific stimulus. While not further explored here, generating an overview of changing pathway dynamics may prove fruitful in understanding how responses to a stimulus are propagated through the cell. This would prove most useful when analyzing how a tumor responds to treatment and what specific pathways function together to mount a response to negate the treatment.

4.2.3 A/B Compartment Switching

Given that changes in nuclear structure can dramatically influence gene expression, we sought to explore if *TCF7L2* silencing alters nuclear structure as a means of regulating gene expression. We found this to be logical as β-catenin, the transcriptional activator of the TCF/LEF transcription factors, recruits a variety of chromatin remodeling factors and may therefore influence the organization of the chromatin surrounding WNT signaling target genes. I performed Hi-C on *TCF7L2*-silenced SW480 cells over the inverted transfection time series. Hi-C libraries were sequenced and analysis was performed by Gabrielle Dotson, a graduate student in the Laboratory of Dr. Indika Rajapakse at the University of Michigan, Ann Arbor. Sequencing resulted in ~150 million total reads per sample and ~125 million uniquely mapped reads. The genome was divided into 100kb bins, which were then partitioned into A/B compartments according to the sign of the chromatin accessibility metric called the Fiedler vector, which is the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix. Positive vector values denote the ‘A’ compartment, which corresponds to euchromatin, and negative vector values denote the ‘B’ compartment, which corresponds to heterochromatin. Gene expression values are shown in blue for chromosomes 13 and 20 at the first and last time points (control and 72 hours post-transfection)(Figure 22). The A/B compartments are shown immediately below in green (euchromatin) or red (heterochromatin). Euchromatin domains correspond closely with gene expression (Figure 22). Chromosomes 13 and 20 were selected as they are commonly gained in colon cancer.

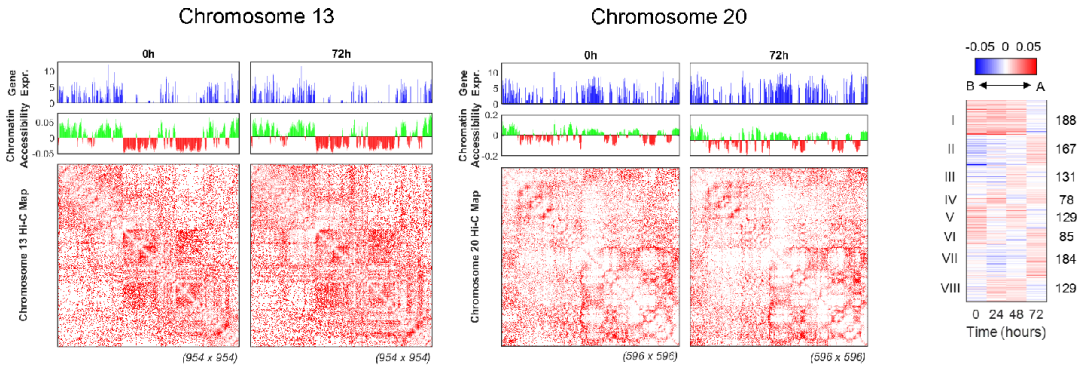

Figure 22 | Influence of A/B compartment partitioning on gene expression. Gene expression for Chromosome 13 (left) and Chromosome 20 (middle) is shown in blue, while A/B compartments are shown in green and red. Contact matrices are shown on the bottom in red. A/B compartment switching dynamics are shown on the right. A/B compartment switching was clustered into eight distinct groups, represented by roman numerals on the left, with the number of genes falling into each specific pattern shown on the right.

To understand A/B compartment dynamics over time, the number of A/B compartment domains were determined for each chromosome at each time point. Overall, ~4% of the genome underwent an A/B compartment switch at each time point, which encompassed 1,305 genes. K-means clustering was used to divide A/B compartment switching into eight patterns, representing either a unidirectional switch (from A to B or B to A) or a bidirectional switch (a permutation of A-to-B-to-A) over the time series (Figure 22). Despite the switching events, the overall ratio of A/B compartments genome-wide remained relatively unchanged with slightly more loci residing in the A compartment (~52%) than in the B compartment (48%). The influence of A/B compartment switching on gene expression was minimal as only 1.3% of the genes within the A/B switching regions underwent a significant change in gene expression.

4.2.4 Local Chromatin Organization and the *CEACAM* Genes

Local partitioning of chromatin into topologically associating domains (TADs) provided insight into the coupling of structure and function. Genes occupying the same TADs have been shown to be co-expressed, likely by accessing the same cluster of transcriptional machinery as a result of their spatial proximity. The number of TADs genome-wide fluctuated by ~5% over the time series. To gain insight regarding which factors may be involved in the TAD domain switching, we used publicly available CTCF and TCF4 ChIP-seq data sets from the HCT116 colorectal cancer cell line. We observed that the binding locations of CTCF in HCT116 matched ~75% of the TAD boundaries of SW480 while the binding locations of TCF4 matched ~41% of the TAD boundaries. However, while TAD boundary binding by CTCF was relatively consistent across chromosomes, the binding of TCF4 fluctuated, with the highest degree of TCF4 binding found on chromosome 19.

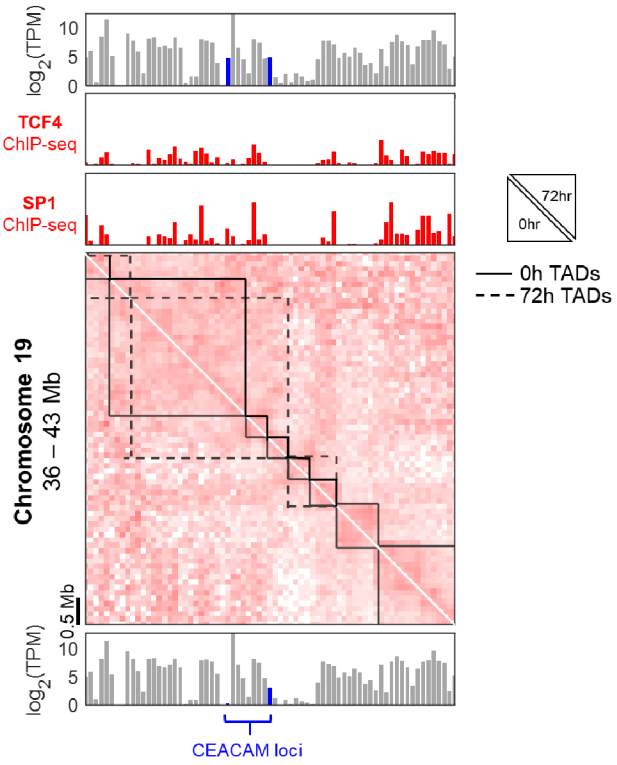

Figure 23 | Expression of the *CEACAM* loci at 0 hours (bottom) and 72 hours (top) with TCF4 and SP1 DNA binding in red and Hi-C contact matrices in pink with calculated TADs denoted by the solid (0 hours) and dotted (72 hours) black lines.

On chromosome 19, an ~10Mb span of the genome houses a group of *CEACAM* family genes, which are involved in colorectal cancer progression and metastasis. Several *CEACAM* genes in this region were potently up-regulated over the time series. At the initial time point, *CEACAM5*, *CEACAM6*, and *CEACAM7* were located within the same TAD, while *CEACAM1* and *CEACAM8* were located in an adjacent, smaller TAD (Figure 23). Changes in expression of the two *CEACAM* gene clusters after 24 hours were negligible. However, after 48 hours, the boundary separating the two TADs was lost, joining *CEACAM5*, *CEACAM6*, and *CEACAM7* into a larger TAD with *CEACAM1* and *CEACAM8*. Combination of the TAD domains as well as a significant increase in expression for both *CEACAM* groups occurred seemingly simultaneously at 48 hours. Gene expression for both groups continued to increase following the TAD boundary loss. The increase in gene expression likely represents increased accessibility to the transcriptional machinery for both gene clusters.

To identify which transcription factors may be mediating this TAD partitioning loss, a transcription factor enrichment analysis was performed using oPOSSUM-3. Over-representation of transcription factor binding sites (TFBS) in the DNA sequences of the 64 genes occupying the differentially conformed region of chromosome 19 were investigated. The top hit, SP1, is found ubiquitously and is involved in chromatin remodeling. To further explore the role of SP1, and TCF4, in mediating this TAD boundary change, we used publicly available SP1 and TCF4 ChIP-seq data from the HCT116 colorectal cancer cell line. We found that both SP1 and TCF4 bind the TAD boundary region observed in SW480 and it is therefore feasible that they are responsible for the repartitioning of this domain (Figure 23).

To determine whether *CEACAM1* up-regulation is prevalent in colon cancer or if it is unique to SW480, we used several colon cancer lines with varying genetic backgrounds which span the major molecular subtypes of colon cancer. SW480 has biallelic mutations in *APC* and is microsatellite stable (MSS), while DLD1 has biallelic *APC* mutations and is microsatellite instable (MSI). HCT116 and LS174T have wild-type *APC* and instead carry activating mutations in *CTNNB1* and are both MSI. COLO201 has biallelic *APC* mutations, is MSS, however was derived from a metastasis, whereas SW480 was derived from a primary tumor. Silencing of *TCF7L2* was performed in each of the cell lines and RNA was extracted, converted to cDNA, and qPCR was performed using primers for *TCF7L2* and *CEACAM1*, and normalized to the expression of the reference gene *YWHAZ*. Silencing of *TCF7L2* resulted in up-regulation of *CEACAM1* in SW480, DLD1, HCT116, and LS174T, demonstrating that the up-regulation of *CEACAM1* is observed across the spectrum of colon cancer lines (Figure 24). The up-regulation is weak in COLO201, though the cause is unknown.

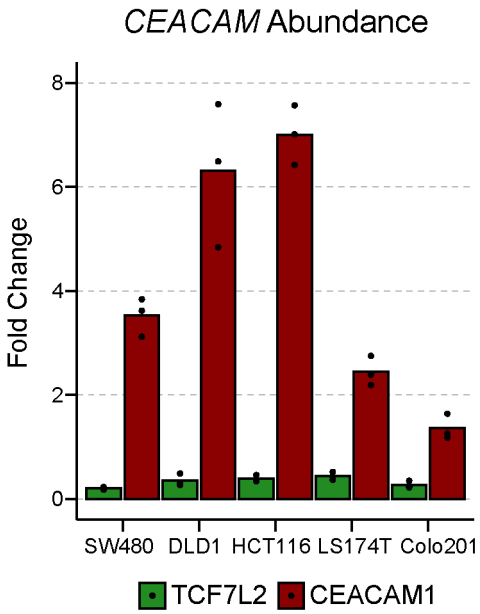

Figure 24 | Up-regulation of *CEACAM1* upon silencing of *TCF7L2* in a select group of colon cancer lines (n=3).

4.2.5 Chromatin Structure and Gene Expression

Following the exploration of the influence of TCF4 at the genome-wide and local levels, we sought to compare how structural organization may influence the expression of genes associated with specific biological functions. We utilized the Frobenius norm, which describes the space of a vector (gene expression) or matrix (chromatin interaction frequencies). We find that, generally, the pathways which undergo more drastic changes in expression originally occupy a more open conformation (Figure 25).

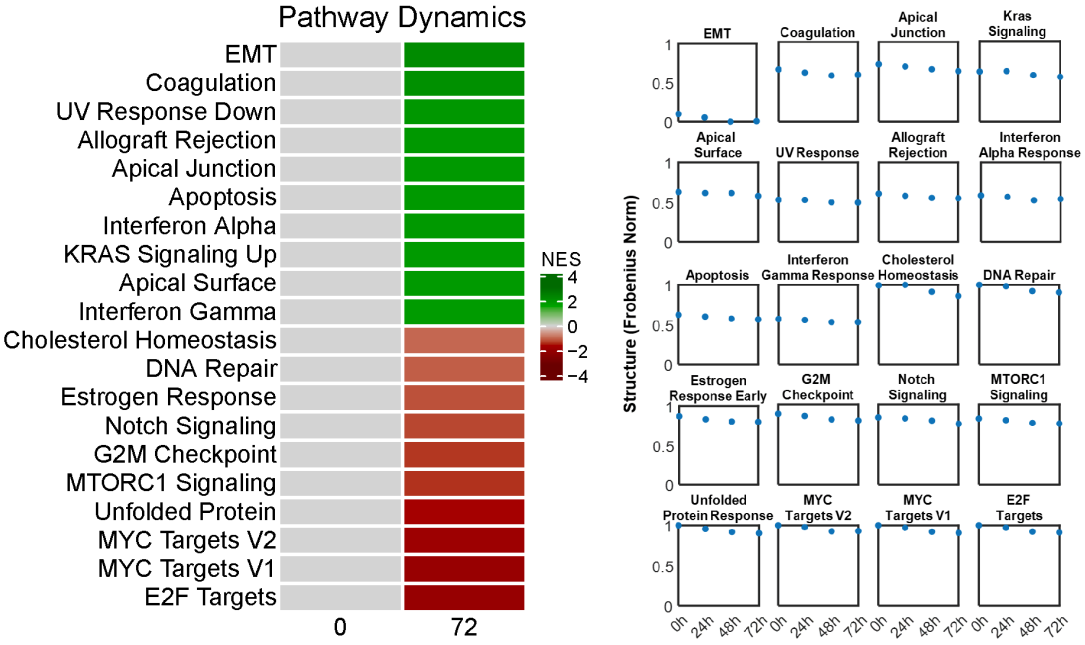

Figure 25 | Gene expression (left) and chromatin structure (right) dynamics for the top ten up-regulated and top ten down-regulated pathway gene sets. Gene expression is ordered by decreasing NES score. Chromatin structure is ordered based on both the Frobenius norm and the order from gene expression.

For instance, strongly down-regulated pathways such as E2F Targets, MYC Targets V1, MYC Targets V2, and Unfolded Protein Response pathway genes reside in a more open conformation than do the moderately down-regulated genes in the MTORC1, Notch, G2M Checkpoint, and Estrogen Response Early pathways. This remains largely true for the up-regulated pathways. However, the exceptions are the EMT, Cholesterol Homeostasis, and DNA Repair pathways. The EMT pathway undergoes the most dramatic up-regulation in expression, yet is in the most constricted conformation, while the Cholesterol Homeostasis and DNA Repair pathways have the least change in expression yet still reside in one of the most open conformations.

According to these results, we find that the conformation of the chromatin is not a necessary determinant of gene expression, but rather, functions as a filter. Genes which reside in open conformations are easily accessed by the transcriptional machinery and are more likely to experience drastic responses in gene expression, while genes in a more constricted environment are less accessible and therefore undergo less dynamic changes in expression. Results from this study, regarding the influence of TCF4 on genome structure, have been published[179].

4.3 A WNT Signaling Intrinsic Feedback Loop: TCF4 and LEF1

4.3.1 *TCF7L2* Silencing and the other WNT Transcription Factors

Upon silencing of *TCF7L2*, I determined the expression of the other WNT signaling factors, as WNT signaling is known to contain several self-regulating feedback loops. I noticed a potent up-regulation of *LEF1*, while expression of *TCF7* and *TCF7L1* remained relatively constant (Figure 26). As previously shown, the expression of *TCF7L1* and *TCF7L2* is high in normal colon while the expression of *TCF7* and *LEF1* is low. In colon cancer, the expression of *TCF7L2* remains high, the expression of *TCF7L1* is lost, while the expression of *TCF7* and *LEF1* is increased. However, does the expression of *TCF7* and *LEF1* have a functional impact or is it a common byproduct of a mutated WNT signaling pathway aimed at increasing nuclear β-catenin levels?

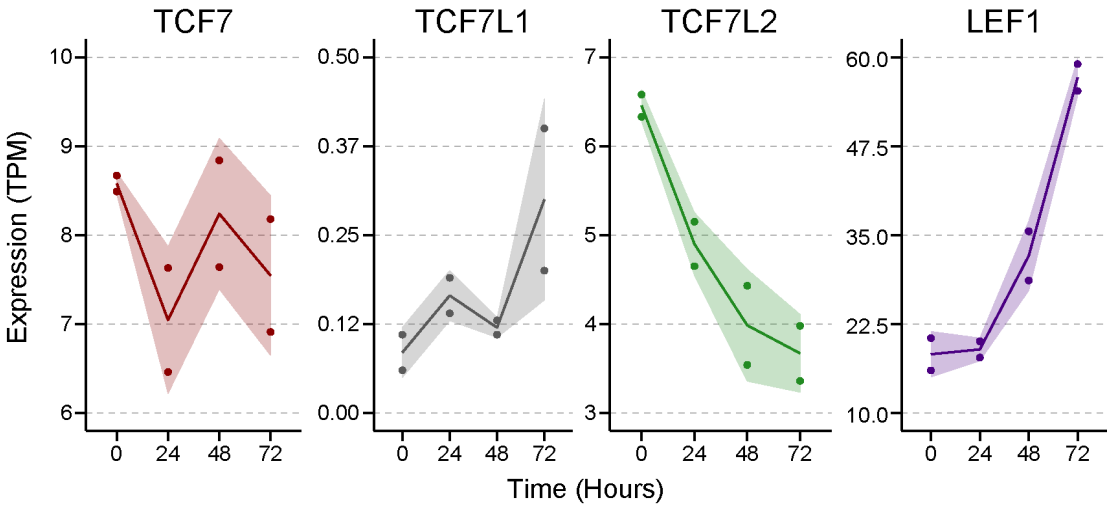

Figure 26 | WNT transcription factor expression upon silencing of *TCF7L2*. The dots represent biological replicates (n=2), the line represents the average expression, and the shaded ribbon represents the standard deviation.

Given the potent induction of *LEF1*, I considered pursuing the function of LEF1, as well as the mechanisms by which TCF4 mediates *LEF1* repression, to be worthwhile. Previous reports in the literature have suggested that the LEF1 isoforms expressed in colon cancer contain the β-catenin binding domain and are therefore considered to be activating. To determine whether the LEF1 isoforms are transcriptionally competent, I performed a western blot to confirm that the up-regulation of *LEF1* results in increased LEF1 (protein) abundance as well as a dual-luciferase reporter assay to determine the impact of *TCF7L2* silencing and *LEF1* over-expression on WNT transcriptional activity.

The western blot for LEF1 up-regulation was performed using the same samples that were used for the western blot in Figure 19. As can be seen, the amount of nuclear LEF1 protein increased progressively over the time series, matching the up-regulation of *LEF1* observed at the RNA level (Figure 27). No LEF1 protein was found in the cytoplasmic protein fraction (results not shown). To determine whether the expressed isoforms of LEF1 were transcriptionally competent, I performed a dual-luciferase assay on *TCF7L2*-silenced SW480 cells. The luminescence from the firefly luciferase, under the control of the TCF/LEF cognate motif, was normalized to the luminescence from the renilla

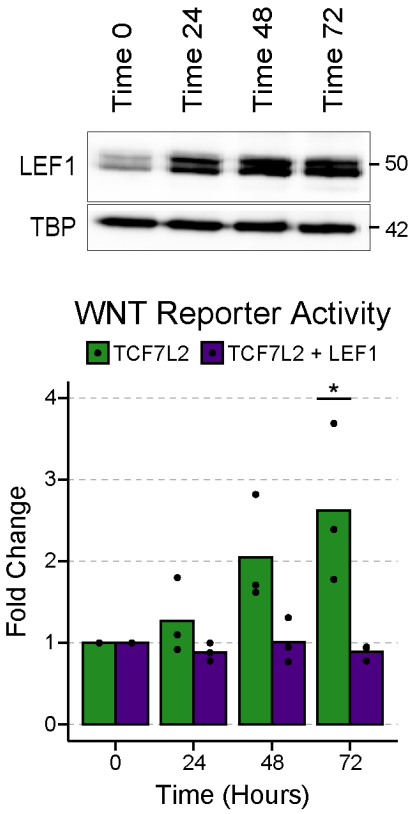

Figure 27 | Western blot for LEF1 upon silencing of *TCF7L2* (top) and WNT Reporter assay (bottom) for silencing of *TCF7L2* alone (n=3) or *TCF7L2* and *LEF1* simultaneously (n=3). The dots represent the replicates, the bars represent the average, and the significance was calculated with Student's two-sided t-test (*$p < 0.05$).

luciferase, which was constitutively active. Negative control samples exhibited basal levels of firefly luciferase expression, while positive control samples were highly luminescent (results not shown). The fold change in expression was determined between the time points (24, 48, and 72) and the control (0 hours). Upon silencing of *TCF7L2*, which results in potent over-expression of *LEF1*, we observe a 2.5-fold increase in reporter activity, shown in green (Figure 27). To determine whether this was a result of *LEF1* up-regulation, I then performed a double-silencing reporter assay, shown in purple, in which *TCF7L2* was silenced (to induce *LEF1* over-expression) and *LEF1* was simultaneously silenced (to suppress the *LEF1* over-expression). The increase in reporter activity caused by silencing *TCF7L2* alone was lost, demonstrating that *LEF1* was responsible for the increase in reporter activity. The expressed LEF1 isoforms are therefore transcriptionally competent.

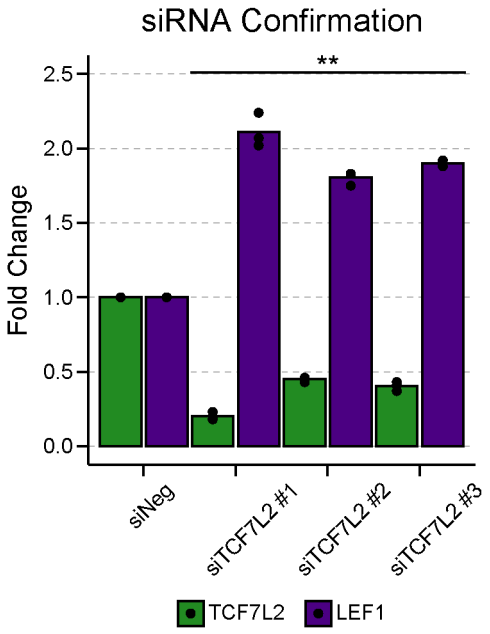

Figure 28 | TCF4-*LEF1* response in SW480 with several siRNAs. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

I then confirmed that the TCF4-*LEF1* response was not a result of siRNA off-target effects by using three different siRNAs, each of which binds a different region of the *TCF7L2* transcript (Figure 28). All three siRNAs resulted in *LEF1* up-regulation, the strength of which depended upon the efficiency of *TCF7L2* silencing, indicating a dose-dependent response. Then, to ensure that the TCF4-*LEF1* response was not unique to SW480, I silenced *TCF7L2* in several colon

cancer cell lines spanning various molecular subtypes of colon cancer. HCT116 and LS174T have wild-type *APC*, while the remaining cell lines have mutated *APC*. LoVo, SW620, and Colo201 are derived from metastatic sites, while HT29, SW480, and DLD1 were extracted from the primary tumor. The TCF4-*LEF1* response was found in SW480 and DLD1, which are the most common colon cancer subtypes (Figure 29).

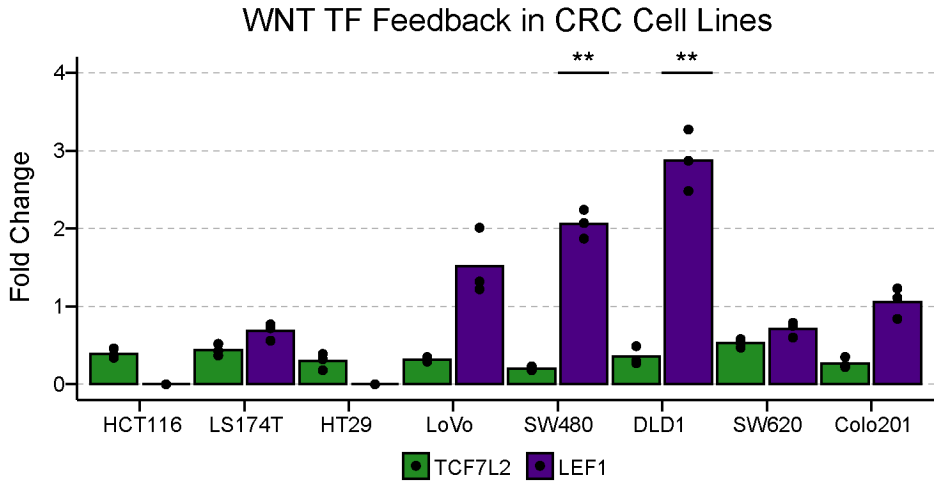

Figure 29 | TCF4-*LEF1* response in several colon cancer cell lines which represent various WNT signaling backgrounds. The dots represent biological replicates (n=3) and the bars represent the average. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

4.3.2 *LEF1* Silencing and the other WNT Transcription Factors

To determine the extent to which LEF1 influences the transcriptome of SW480, I applied for a small NIH internal grant (STARS award) to perform another round of RNA sequencing. The grant was awarded and a new set of transfections were performed for *LEF1* silencing, using the same procedure used to silence *TCF7L2*. Silencing of *LEF1* had previously been confirmed at the RNA and protein levels (Figure 30). *LEF1* transcript abundance decreased by 2-fold 48 hours post-transfection and 4-fold by 72 hours post-transfection. Abundance of LEF1 protein remained unchanged at 24 hours, decreased ~75% by 48 hours, and decreased ~97% by 72 hours post-transfection.

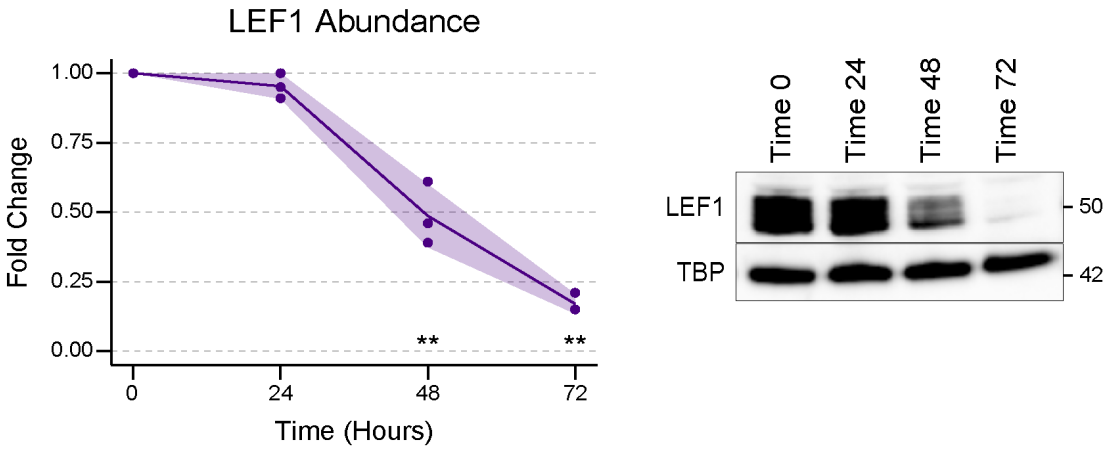

Figure 30 | *LEF1* silencing in the SW480 cell line. *LEF1* transcript abundance (left) was determined using qPCR and normalized against the *YWHAZ* reference gene. The dots represent biological replicates (n=3), the line represents the average, and the shaded purple ribbon represents the standard deviation. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$). LEF1 protein abundance was determined using western blot (right) with TBP as loading control.

Following successful silencing of *LEF1*, RNA sequencing was performed. RNA was harvested from *LEF1*-silenced SW480 cells and treated with DNase I to degrade potential contaminating genomic DNA. RNA integrity was determined by the CCR Genomics Core using an Agilent TapeStation 4200. Samples with a RIN of 9 or higher were sent to the CCR Sequencing Facility for library preparation and sequencing.

The expression of the WNT signaling transcription factors was plotted to determine their response to *LEF1* silencing (Figure 31). The expression of *LEF1* was potently down-regulated, confirming successful silencing, while the response of *TCF7* and *TCF7L1*, similar to their response to *TCF7L2* silencing, did not demonstrate a clear trend. *TCF7L2* demonstrated a slight increase at the last time point, though no clear trend was observed for *TCF7L2* (Figure 31). The silencing of *LEF1* therefore leaves the expression of the remaining TCF/LEF factors largely unperturbed. The interaction

between *LEF1* and *TCF7L2* is therefore not reciprocal, as *TCF7L2* silencing results in a clear up-regulation of *LEF1*, while the silencing of *LEF1* does not result in a clear up-regulation of *TCF7L2*. The up-regulation of *LEF1* upon silencing of *TCF7L2* suggests that TCF4 has repressive capabilities which LEF1 lacks.

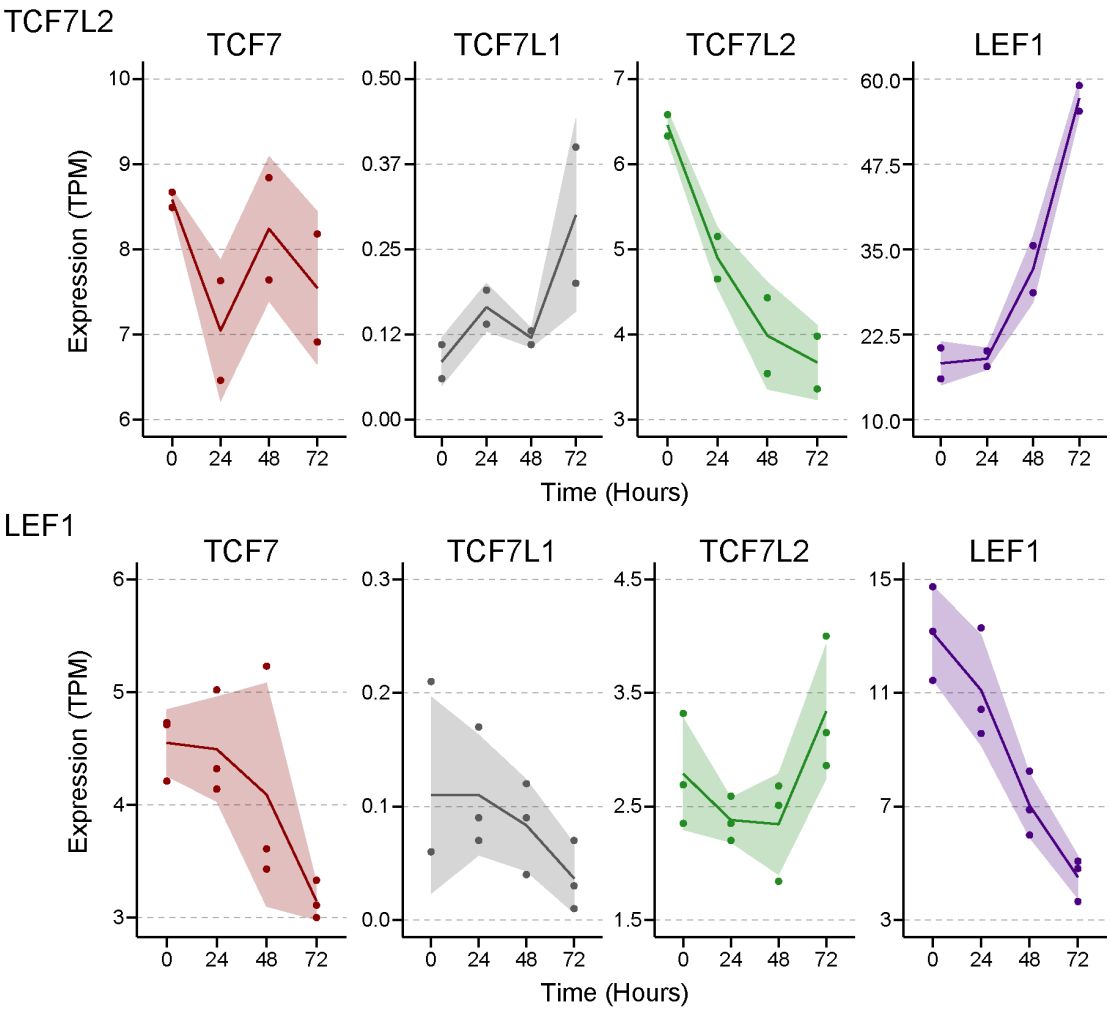

Figure 31 | Expression of the WNT transcription factors in *TCF7L2*-silenced (top) or *LEF1*-silenced (bottom) SW480 cells. Two biological replicates (n=2) are plotted for *TCF7L2* silencing with the solid line representing the average. The shaded ribbon denotes the standard deviation. Three biological replicates (n=3) are plotted for *LEF1* silencing with the solid line representing the average and the shaded ribbon denoting the standard deviation.

After determining the response of the WNT signaling transcription factors to *LEF1* silencing, I explored the extent to which LEF1 influences the transcriptome. Volcano plots were used to visualize the results of differential gene expression analysis upon *LEF1* silencing (Figure 32). The most dramatic changes in gene expression were again observed by the last time point, in accordance with the greatest reduction of LEF1 protein seen by western blot and silencing of *LEF1* transcripts observed with qPCR.

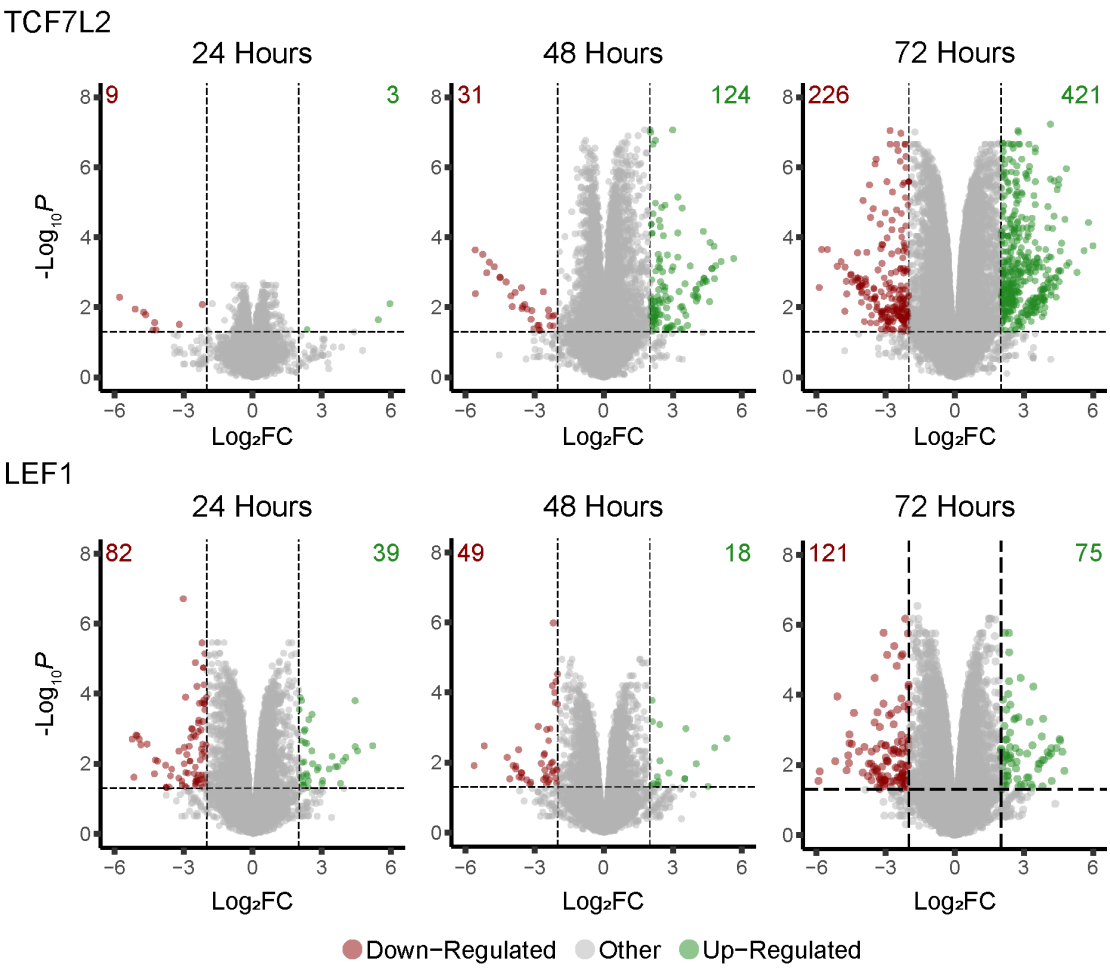

Figure 32 | Volcano plots for differential gene expression upon silencing of *TCF7L2* (top) or *LEF1* (bottom). Two biological replicates were used for *TCF7L2* and *LEF1* (n=2). The threshold for $\log_2$ fold change was set to two and the significance was set to 0.05 (FDR adjusted *p*-value). Genes which exceeded the fold change threshold and were lower than the adjusted *p*-value were shaded green for up-regulated and red for down-regulated.

In contrast to the response seen upon silencing *TCF7L2*, wherein the majority of genes are up-regulated, the response to *LEF1* silencing consisted of a general down-regulation of gene expression. Put another way, an up-regulation is observed when *LEF1* is over-expressed, as a result of silencing *TCF7L2*, and a down-regulation is observed when *LEF1* is silenced. Given that *LEF1* is the highest expressed WNT transcription factor in SW480, it may not be surprising that the general gene expression pattern follows the expression of *LEF1* closer than *TCF7L2*.

A second distinction between silencing *TCF7L2* and *LEF1* is the pattern of the response. When silencing *TCF7L2*, changes in gene expression occur incrementally, reaching a maximum by the last time point. While *LEF1* silencing also reaches a maximum at the last time point, the 24-hour time point shows more up- and down-regulated genes than the 48-hour time point. This is counterintuitive since the levels of *LEF1* appear unchanged at both the RNA and protein levels as shown by qPCR and western blot at the 24-hour time point. Closer inspection of the up- and down-regulated genes from the 24-hour time point shows that they are similar to the genes from the 72-hour time point, with several immune system genes highly down-regulated, suggestive of biological function (*IL32*, *CD22*, *CD44*). This is unlikely to be the result of a technical issue as the same procedure was used to generate the *TCF7L2*-silenced data. Lipofectamine toxicity is also unlikely as the 72-hour time point and the negative control (0 hours) were in lipofectamine reagent for longer periods of time than the 24-hour time point. One explanation is that the transfection for the LEF1 protein used in the western blot may have been less efficient at the 24-hour time point, leading to higher levels of LEF1 than were observed in the RNA sequencing data.

4.3.3 Transcriptional Dynamics of LEF1

After learning that *LEF1* silencing influences the transcriptome of SW480, and results in a different outcome than silencing *TCF7L2*, I wanted to explore which pathways were affected by silencing of *LEF1* and how this response differed from *TCF7L2* silencing. To this end, I used GSEA and determined the activity of the pathways identified previously, from *TCF7L2* silencing, upon *LEF1* silencing. The goal was to identify how the most differentially expressed pathways under *TCF7L2* silencing (when *LEF1* was over-expressed) behaved when *LEF1* was silenced. The top ten up-regulated and top ten down-regulated pathways from *TCF7L2* silencing were taken and their NESs were calculated using the *LEF1*-silenced RNA sequencing data.

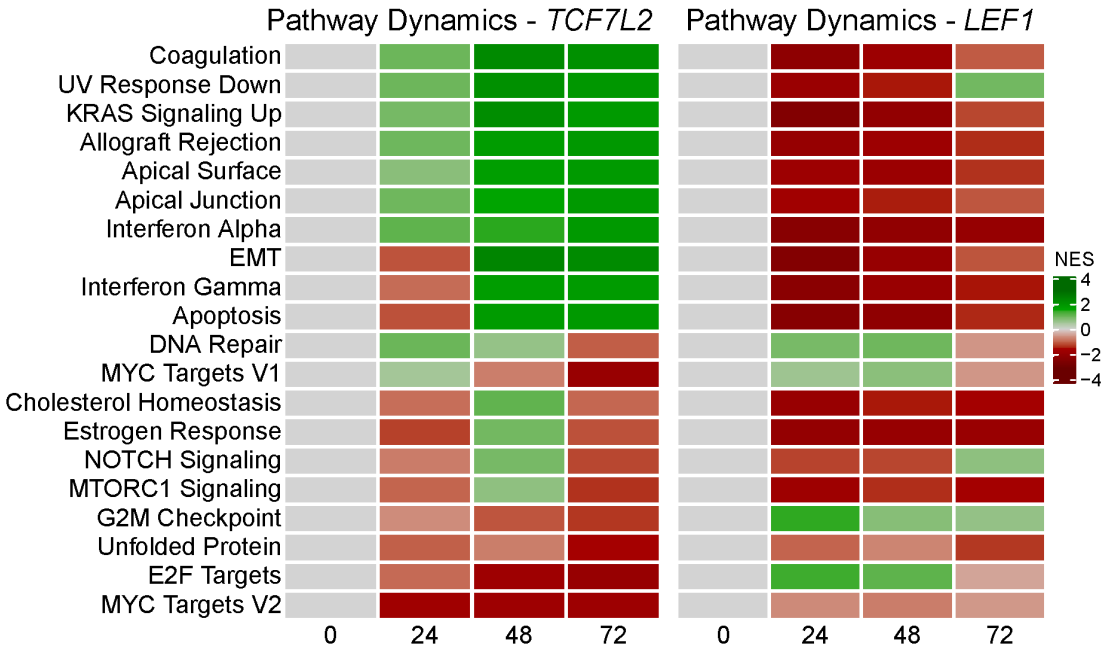

Figure 33 | GSEA of *TCF7L2*- and *LEF1*-silenced SW480 cells. The top ten up-regulated and top ten down-regulated pathways upon *TCF7L2* silencing were plotted (left) and their activity upon *LEF1* silencing was determined and plotted (right).

The results for *TCF7L2* silencing are shown on the left, while the results for *LEF1* silencing, are shown in the same order, on the right (Figure 33). The difference in pathway dynamics is quite striking, as nine of the top ten up-regulated pathways upon silencing of *TCF7L2* were down-regulated upon silencing of *LEF1* by the last time point. This demonstrates that LEF1 plays a central role in regulating the transcriptional program of SW480 cells. The inverse was then performed, by determining the top ten up-regulated and top ten down-regulated pathways upon silencing of *LEF1* and measuring the behavior of these pathways in the *TCF7L2*-silenced data (Figure 34). The top ten down-regulated and top nine up-regulated (there were only nine up-regulated pathways) based on the 72-hour time point, were determined and are shown below on the left. The majority of down-regulated pathways were involved in the immune system, such as IL6, IL2, Interferon Alpha, and Inflammatory Response.

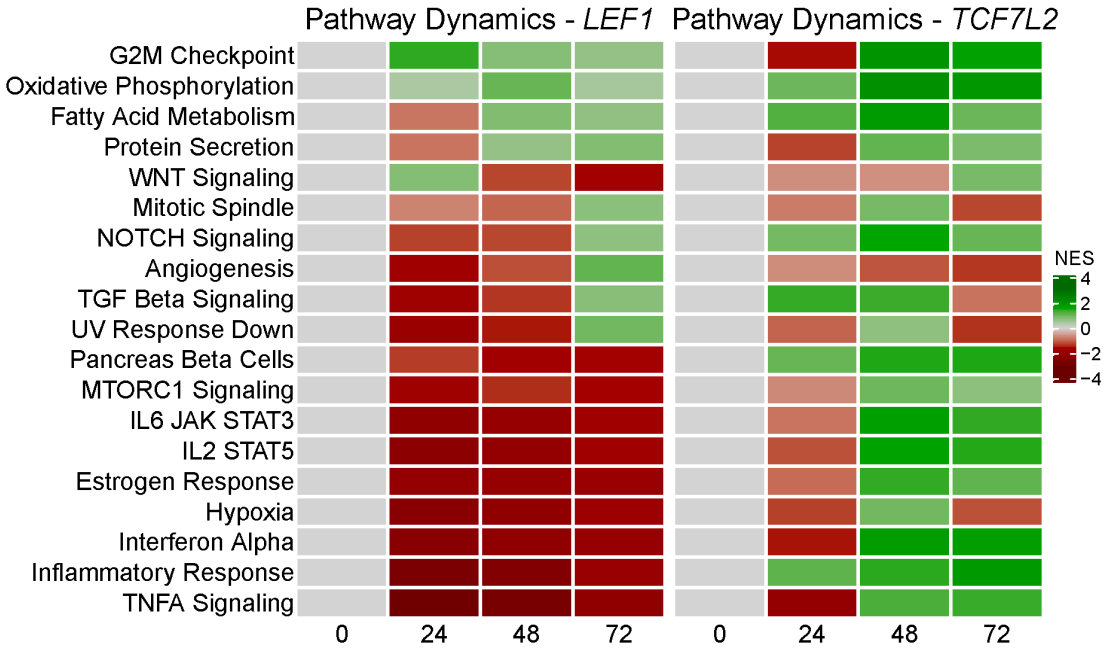

Figure 34 | GSEA for *LEF1*- and *TCF7L2*-silenced SW480 cells. The ten most down-regulated and nine most up-regulated pathways in *LEF1*-silenced cells were determined and plotted (left). The activity of these pathways was then determined in the *TCF7L2*-silenced data and plotted (right).

These results are anticipated given that LEF1 is involved in lymphocyte development and drives the expression of the TCRα enhancer in lymphocytes. Only a handful of pathways were up-regulated at more than one time point. When the activity of the most dynamically changing pathways in the *LEF1*-silenced cells are determined in the *TCF7L2*-silenced cells, we see that the majority of down-regulated pathways exhibit the opposite response and are up-regulated. This demonstrates the importance of LEF1 as a central driver of transcription in SW480.

4.3.4 TCF4 and LEF1 Target Genes

After having learned that expressed LEF1 isoforms are transcriptionally competent and that LEF1 plays a noticeable role in mediating transcription in SW480 cells, the next goal was to generate a list of TCF4 and LEF1 specific target genes to better understand how these transcription factors coordinate transcription. I decided the best method to accomplish this was Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) on TCF4 and LEF1. The previously described STARS award covered the costs of RNA sequencing as well as ChIP sequencing. The SimpleChIP Kit from Cell Signaling Technologies and ChIP-grade, validated antibodies for TCF4 and LEF1 were used. ChIP DNA was sent to the CCR Genomics Core for quality control using an Agilent DNA TapeStation and sent to the CCR Sequencing Facility for library preparation and sequencing.

Sequencing results were collected from the sequencing facility and the CCBR Pipeliner for ChIP-seq was used to process the samples. A list of TCF4 and LEF1 binding sites was used to generate target gene lists for TCF4 and LEF1. The gene lists for both transcription factors were large and it was difficult to define whether genes represented true binding sites in which TCF4 or LEF1 could have a

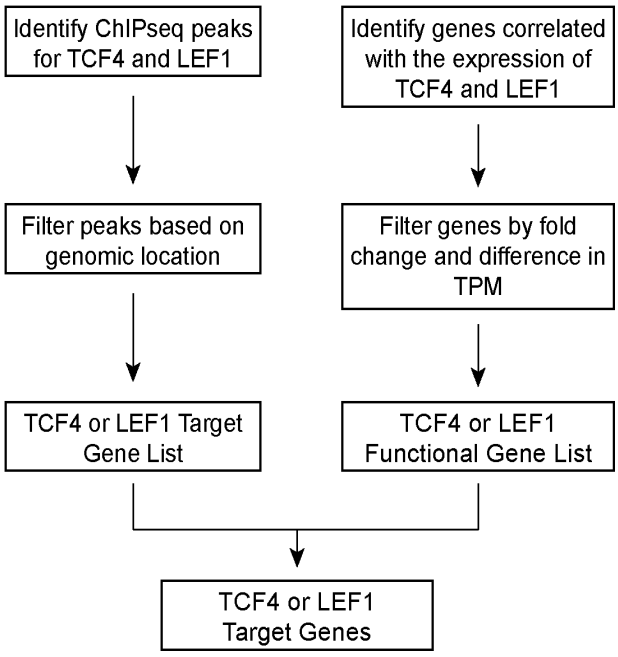

Figure 35 | Filtering strategy for identifying TCF4 and LEF1 target genes.

clear impact on gene expression. To navigate this, I designed a series of steps to generate a biologically relevant gene list combining RNA sequencing as well as ChIP-seq data (Figure 35). The process began by generating a list of all peaks for TCF4 and LEF1 and filtering based on the location of the peak in relation to the gene. Peaks which were in the promoter region or upstream of the gene were selected, while genes residing within the body of the gene where discarded. A different list was then generated by correlating the expression of a gene to the expression of *TCF7L2* or *LEF1*. If TCF4 or LEF1 have biologically relevant control over target gene expression, then their expression should be correlated (Pearson correlation >0.7 or < -0.7). Genes with low change in TPM or low fold change between the first and last time points were excluded as they are likely not biologically relevant. Genes occurring in both the ChIP-seq based list and the RNA sequencing based list were considered 'true' target genes.

Gene	TF Correlation	TPM Shift	Log_2FC	TF Binding
ENC1	-0.81	32.97	3.72	TCF4
CD44	-0.77	104.73	3.41	TCF4
LEF1	-0.77	38.90	1.64	TCF4
MYC	0.93	-190.65	-0.89	TCF4
CCND1	0.85	-65.53	-0.92	TCF4
AXIN1	0.89	-10.14	-0.91	TCF4
MYC	0.47	-28.38	-0.36	LEF1
CCND2	0.64	-33.29	-0.33	LEF1

Table 5 | TCF4 and LEF1 target gene list excerpts. The dotted line distinguishes the TCF4 from the LEF1 target genes with *MYC* being bound by both TCF4 and LEF1.

Analyzing the number of true target genes for TCF4 resulted in a list of 384 genes, an excerpt of which is shown in Table 5. The genes consisted mainly of known WNT signaling target genes including *MYC*, *CCND1*, *AXIN1*, *ENC1*, and *CD44*. The list also included *LEF1*. When performing the same procedure for LEF1, no true target genes were identified. The procedure was performed again with less strict requirements (correlation of 0.45 instead of 0.7). This resulted in 12 true target genes for LEF1. Two of the twelve target genes are in the WNT signaling pathway, *MYC* and *CCND2* (Table 5). Given that *MYC* is bound by both TCF4 and LEF1 suggests that over-expression of *LEF1* in SW480 may benefit the colon cancer cell by facilitating the expression of *MYC*. However, given the strong effects of LEF1 on transcriptional patterns shown by GSEA, the small true target gene list is unexpected. A likely explanation for this is that *LEF1* is not a transcription factor normally expressed in colon cells, as it is not detected in normal colon tissue. Given that all four TCF/LEF factors can bind the same WNT motif, due to their highly conserved HMG DNA binding domain, it is likely that accessory proteins exist which facilitate the binding of WNT transcription factors to their target genes. These accessory proteins are likely tissue specific and therefore associate with the WNT/LEF factor native to that tissue, as the amino acid sequence outside of the conserved HMG domain diverges between the different WNT transcription factors. This could explain the large number of TCF4 true target genes identified here. Given that LEF1 would not be able to bind these tissue specific accessory proteins, its ability to influence transcription would therefore be based on indirect interactions, which may facilitate the actions of other transcription factors, or the binding sites of LEF1 may lie outside of the promoter region, perhaps in distant enhancer regions, and therefore may have escaped detection in this analysis.

4.3.5 The Capability of TCF4 to Regulate the Expression of *LEF1*

Thus far I have determined that *LEF1* is highly expressed in SW480 colon cancer cells and *LEF1* expression increases dramatically upon silencing of *TCF7L2*. The expressed variants of *LEF1* are transcriptionally competent and these expressed variants are capable of dramatically changing the transcriptional landscape of SW480, which is accomplished by indirect as opposed to direct binding of LEF1 to its target genes. I then explored how TCF4 mediates repression of *LEF1*. Given that *LEF1* appeared as one of TCF4's true target genes, the interaction is likely direct. Indeed, TCF4 shows a clear peak at the *LEF1* promoter in SW480 cells (Figure 36).

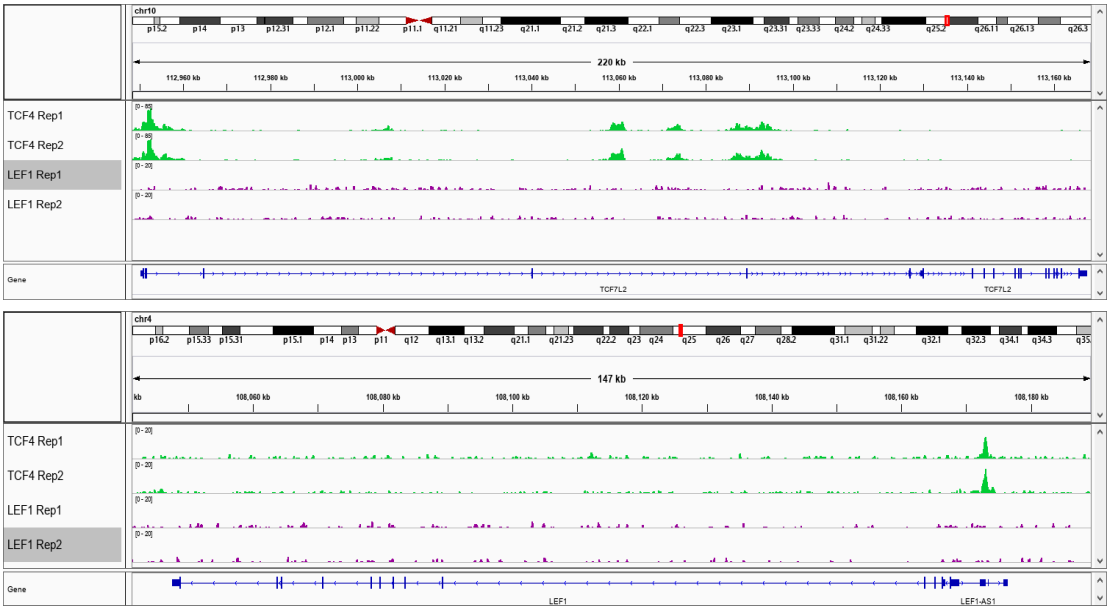

Figure 36 | TCF4 and LEF1 binding at the *TCF7L2* (top) and *LEF1* (bottom) loci in SW480 colon cancer cells. Note that *LEF1* is located on the reverse strand, going from right to left.

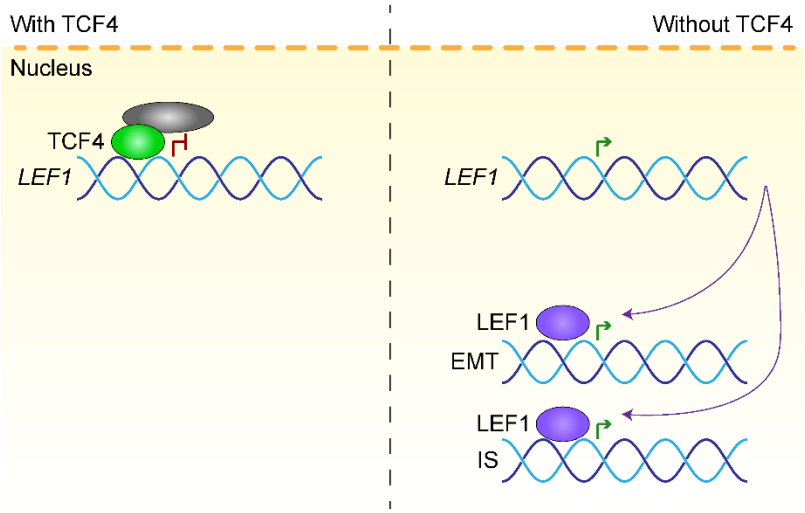

Figure 37 | Diagram depicting the TCF4-*LEF1* relationship based on the data collected.

In SW480 cells, TCF4 is present and can bind the *LEF1* promoter region. Silencing of *TCF7L2*, which greatly reduces TCF4 abundance results in a potent up-regulation of *LEF1*. These LEF1 isoforms are transcriptionally competent and lead to the activation of genes involved in EMT as well as various Immune System (IS) associated pathways (Figure 37). Given the direct interaction between TCF4 and *LEF1*, as well as the potent up-regulation of *LEF1* expression upon TCF4 loss, suggests that TCF4 represses *LEF1* expression. However, TCF4 is unable to influence transcription alone, so the question remains, which binding partner of TCF4 conveys the repressive function. The isoforms of TCF4 expressed in colon tissue are capable of repressing gene expression by two methods. The first method is by binding of a TLE transcriptional repressor to TCF4, which associates with histone deacetylases to repress gene transcription. The second method relies on the binding of a CtBP transcriptional repressor to TCF4.

I first pursued the capability of the TLE factors to mediate TCF4-based repression since they are much better understood than the CtBP factors, and because several studies present conflicting evidence regarding the role of the CtBP factors in transcription. Influencing TLE activity can be initiated by several methods. The first approach is plasmid-based over-expression of TLE. This is unlikely to produce the intended effect since β-catenin has been reported to displace the TLE factors while binding to TCF4. Over-expression of TLE, in a colon cancer setting with abundant nuclear β-catenin, would likely be ineffective due to the competition between β-catenin and TLE. A second method is silencing of β-catenin to allow endogenous levels of TLE to resume their repressive function alongside TCF4 with decreased levels of β-catenin. The limitation of this approach is the effectiveness of the siRNA-mediated silencing of β-catenin as well as the levels of TLE necessary to influence a repressive effect. The third option was to reintroduce wild-type *APC* cDNA, via a plasmid, into SW480 cells, which have biallelic *APC* mutations. Given that the majority of colon cancers carry biallelic *APC* mutations, reintroduction of *APC* is likely the best method to constrain WNT signaling activity and reestablish the repressive effects of TCF4.

A plasmid containing the cDNA for wild-type *APC* was purchased from Addgene. APC is a large protein (312kDa) and contains a variety of motifs. The domains for binding β-catenin and AXIN, a member of the destruction complex, lie closer to the C-terminus, and this region is the hotspot for mutations in colon cancer. Previous work in the inherited colorectal cancer syndrome, familial adenomatous polyposis, has demonstrated that the severity of polyposis depends on the location of the *APC* mutation, and its proximity to the mutation hotspot region, referred to as the Mutation

Cluster Region (MCR). The exact reason for the severe phenotype is not completely known, though β-catenin levels play a role. I hypothesized that, since *APC* mutations influence β-catenin levels, which may then influence the ability of TCF4 to repress *LEF1*, there may be a connection from APC to *LEF1*. I knew that the location of the *APC* mutation would influence β-catenin levels, so I hypothesized that down-regulating WNT signaling may influence the ability of TCF4 to mediate repression of *LEF1*. To test this hypothesis, I generated three truncated APC proteins of different lengths (60, 70, and 80kDa) to mimic the different mutations in FAP and colon cancer.

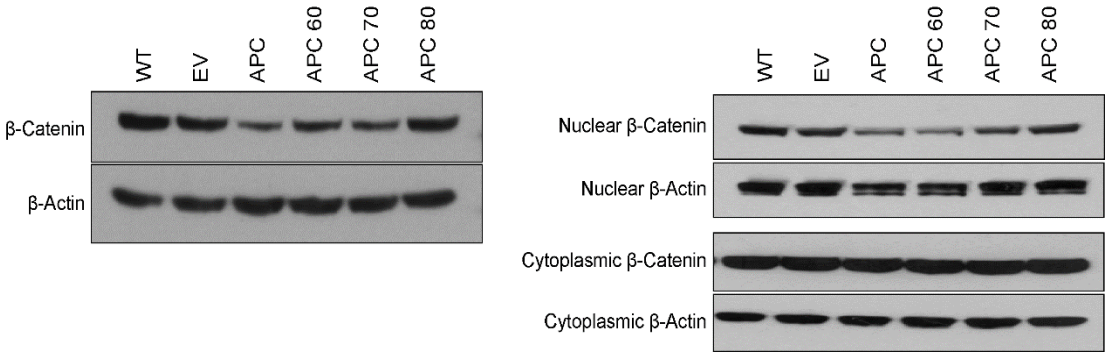

Figure 38 | SW480 cells were transfected with the EV, APC, APC 60, APC 70, or APC 80 plasmid. Total protein was harvested and a western blot was performed for total β-catenin (left). Nuclear and cytoplasmic protein extractions followed by western blot for active β-catenin (right).

The APC, APC 60, and APC 70 plasmids all demonstrated relatively equal expression of *APC*, or of the appropriate fragment using qPCR (data not shown). The expression of APC 80 was ~50% lower, though the plasmid was used in experiments. Western blot could not be performed because no antibodies targeting the region of APC found in the fragments could be found. Additionally, efforts to generate a custom antibody with Rockland Immunochemicals, Inc. for the APC fragments proved unsuccessful.

I proceeded to validate that reintroduction of APC had the intended effects on SW480. Total as well as nuclear and cytoplasmic protein fractions were harvested and a western blot was performed to determine β-catenin down-regulation. Reintroduction of APC, APC 60, and APC 70 decreased the amount of total β-catenin (left), while EV and APC 80 did not change total β-catenin levels (Figure 38). APC, APC 60, and APC 70 also decreased nuclear β-catenin levels (Figure 38). EV and APC 80 samples demonstrated no decrease in nuclear β-catenin levels, similar to total β-catenin levels.

None of the plasmids reduced the amount of cytoplasmic β-catenin, demonstrating that the reduction in total β-catenin (seen left), is due to loss of nuclear β-catenin (seen right). To confirm that reduction in nuclear β-catenin resulted in a reduction in WNT signaling activity, a dual-luciferase assay was performed. Results from the WNT reporter assay show that APC, APC 70, and APC 60 exhibit the lowest WNT signaling activity, in line with the reduction in nuclear β-catenin protein levels (Figure 39). APC 80 also demonstrates a reduction in WNT signaling activity, which comes as a surprise given that no reduction in β-catenin levels were observed in the western blots.

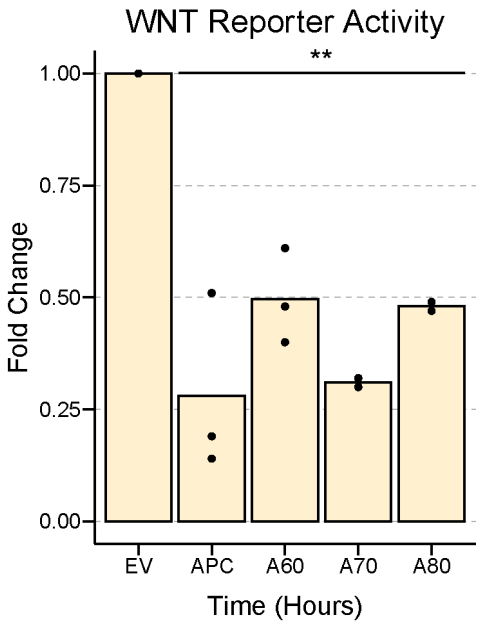

Figure 39 | Dual-luciferase assay was performed on SW480 cells transfected with EV, APC, APC 60, APC 70, or APC 80. Three biological replicates (n=3) are plotted for EV, APC, and APC 60, while two biological replicates (n=2) are plotted for APC 70 and APC 80. Results are normalized to EV. Significance was calculated using Student's two-sided t-test (*p < 0.05, **p < 0.01).

Given that nuclear β-catenin levels are decreased in APC, APC 60, and APC 70 transfected cells, as well as WNT signaling activity as measured with a WNT reporter assay, I performed a growth assay to confirm that reintroduction of APC or the APC fragments reduced the growth rate of SW480 cells (data not shown). Results from the growth assay confirmed the results from the nuclear western blot as well as the dual-luciferase assay. Namely, the plasmids with the slowest growth were APC and APC 70, followed by APC 60 and APC 80. Cells transfected with EV grew the fastest.

After confirming the functionality of the *APC* plasmids, I then determined the expression of *LEF1* in SW480 cells transfected with the various *APC* plasmids (Figure 40). Introduction of full-length APC into SW480 cells decreased the expression of *LEF1* by ~40% which was similar to the effect of APC 60. APC 70 and APC 80 decreased the expression of *LEF1* by ~25%. I can conclude that reintroduction of APC, and fragments thereof, influence the expression of *LEF1*. However, the data is not sufficiently robust to determine whether there is a difference in *LEF1* expression when different regions of APC are present.

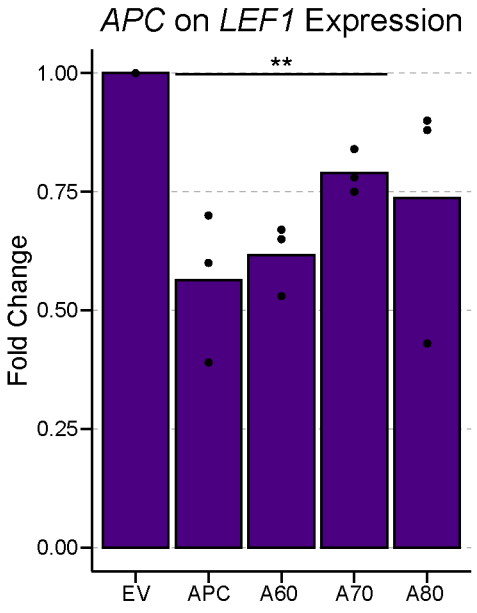

Figure 40 | *LEF1* expression in *APC* plasmid transfected SW480 cells. Three biological replicates (n=3) are plotted per sample. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

I then traveled to the Laboratory of Dr. Michael Boutros at the Deutsche Krebsforschungszentrum (DKFZ) in Heidelberg, Germany, to further explore the role of APC in regulating the feedback interaction between TCF4 and *LEF1*. The Boutros Laboratory has developed several *APC* model cell lines using colon cancer cell lines with wild-type *APC*. The wild-type *APC* were targeted with CRISPR, resulting in biallelic *APC* mutations. The mutations occurred in the MCR region of *APC*, therefore representing clinically relevant mutations. My goal for visiting the laboratory was to determine if APC influences the interaction between TCF4 and *LEF1*.

I used two colon cancer cell lines with wild-type *APC*, HCT116 and RKO, as well as two clones from each cell line, which had biallelic mutations in *APC* artificially introduced using CRISPR. I silenced *TCF7L2* in these cells and performed qPCR to determine whether the TCF4-*LEF1* interaction was altered between the wild-type HCT116 and RKO lines and their *APC* mutated counterparts (Figure 41). I found mutation of *APC* had no impact on expression of *LEF1* in HCT116 cells (left), while in wild-type RKO, silencing of *TCF7L2* lead to an up-regulation of *LEF1*. In the *APC* mutated RKO clones, this relationship was maintained and the levels of *LEF1* up regulation were similar to that of wild-type RKO. These results show that *APC* mutation alone is not sufficient for influencing the TCF4-*LEF1* feedback interaction.

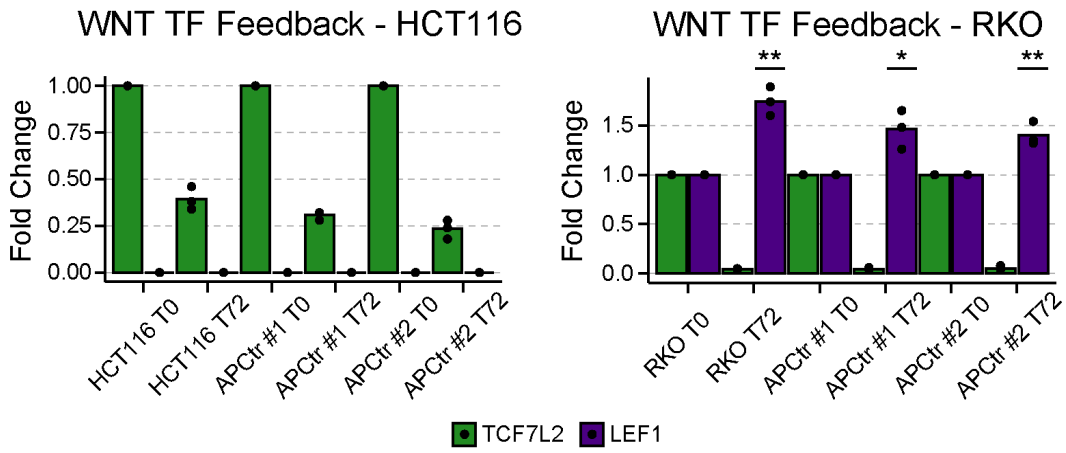

Figure 41 | *TCF7L2* silencing in HCT116, RKO, and two clones with biallelically mutated *APC* from HCT116 and RKO. Transcript abundance of *TCF7L2* and *LEF1* was determined using qPCR. Three biological replicates (n=3) are plotted with *TCF7L2* abundance shown in green and *LEF1* transcript abundance shown in purple. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

The other group of transcriptional repressors known to bind TCF4 are the CtBP family of proteins. In humans, this family consists of two members, CtBP1 and CtBP2. Conflicting roles for the CtBP proteins alongside TCF4 have been described, with some reports indicating a repressive function, while other reports describe an activating function for the CtBP proteins when complexed with β-catenin and TCF4, which occurs in colon cancer. To investigate the effects of the CtBP1 and CtBP2 proteins in mediating the TCF4-*LEF1* interaction, I used siRNAs to target the expression of *CtBP1* and *CtBP2*. Western blot and qPCR were used to assess CtBP1, CtBP2, TCF4, LEF1, and TBP protein levels as well as *TCF7L2* and *LEF1* abundance.

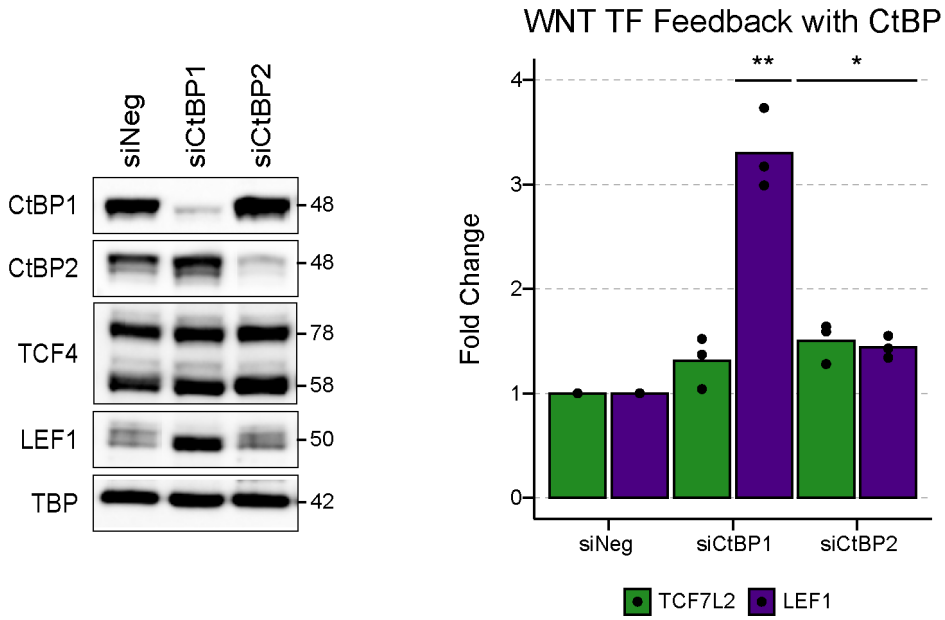

Figure 42 | Western blot (left) showing protein abundance of CtBP1, CtBP2, TCF4, LEF1, and TBP upon treatment with siNeg, siCtBP1, or siCtBP2. Transcript abundance, assayed via qPCR (right) of *TCF7L2* or *LEF1* following exposure to siNeg, siCtBP1, or SiCtBP2. Significance was calculated using Student's two-sided t-test (*$p < 0.05$, **$p < 0.01$).

Silencing of *CtBP1* or *CtBP2* led to a clear reduction of CtBP1 or CtBP2 protein levels, shown by western blot (Figure 42). Loss of CtBP1 or CtBP2 did not influence TCF4 abundance, removing a possible confounding factor from interpreting LEF1 levels. Loss of CtBP1, but not CtBP2, resulted in a potent increase in LEF1 protein abundance, supporting the observation that the CtBP family members are transcriptional repressors. Results from qPCR confirmed the results from western blot.

Chapter 5: Conclusions and Future Directions

5.1.1 Consequences of *TCF7L2* Silencing in Colon Cancer

I studied the consequences of silencing *TCF7L2*, which encodes for TCF4, the major WNT signaling transcription factor in the colon. In colon cancer, activating mutations in the WNT signaling pathway, most often in *APC*, result in the nuclear localization of β-catenin and constitutive WNT pathway transcription mediated by β-catenin/TCF4 complexes. This leads to high expression of *MYC* and *CCND1* as well as other growth promoting genes. As a result of this, TCF4 is generally recognized to be a transcriptional activator due to its association with β-catenin in colon cancer.

Silencing of *TCF7L2* followed by RNA sequencing led to a disproportionate up-regulation in gene expression suggesting that *TCF7L2* may play a repressive, as well as an activating, role in colon cancer. Hi-C experiments on *TCF7L2*-silenced SW480 cells resulted in ~4% of the genome undergoing an A/B compartment switch at each time point, however, the overall ratio of A/B compartments remained constant over the time series. A/B compartment switching did not have a significant influence on gene expression as only 1.3% of genes falling within A/B compartment switching regions also demonstrated a significant change in gene expression. Changes in local chromatin structure did, however, result in a noticeable up-regulation in two groups of *CEACAM* genes on chromosome 19. Over the time series, the two groups, originally located in different TADs, became combined in the same TAD and simultaneously underwent a significant increase in gene expression. TAD joining brought these two related gene groups in close physical space, allowing these previously separated genes to share the same transcriptional machinery, resulting in their up-regulation.

5.1.2 Consequences of *LEF1* Silencing in Colon Cancer

Silencing of *TCF7L2* also resulted in the potent up-regulation of *LEF1*, one of the other WNT signaling transcription factors. LEF1 isoforms were capable of driving WNT signaling activity in a WNT reporter assay, and over-compensated for the loss of TCF4, in terms of overall WNT signaling activity. Concomitant silencing of *TCF7L2* and *LEF1* abrogated the increase in WNT signaling activity, demonstrating that LEF1 was responsible for the increase. Upon silencing of *TCF7L2*, which results in a potent up-regulation of *LEF1*, the expression of differentially expressed genes was disproportionately increased, while when *LEF1* was silenced, the expression of differentially expressed genes disproportionally decreased. Taken together, these results demonstrate that LEF1 is a central regulator of transcription in SW480 cells. GSEA demonstrated down-regulation of MYC and G2M checkpoint pathways upon silencing of *TCF7L2*, in accord with the role of TCF4 as a regulator of MYC expression and cell cycle progression. Silencing of *LEF1* resulted in the down-regulation of immune system centric pathways including IL2, IL6, as well as genes involved in the inflammatory response, in accord with the role of LEF1 as a regulator of the immune system. Comparing GSEA results between *TCF7L2*- and *LEF1*-silenced SW480 cells demonstrated that the most significantly up- and down-regulated pathways follow the expression pattern of *LEF1*. ChIP sequencing was then used to identify the target genes of TCF4 and LEF1. An algorithm was generated which identified true TCF4 and LEF1 target genes, by integrating ChIP-seq data, which demonstrated gene binding, with RNA sequencing data, which demonstrated biological relevance. Only a single gene was found to be overlapping between the two transcription factors, which was *MYC*. This shows that LEF1 in colon cancer can help TCF4 drive the expression of

MYC, however otherwise, LEF1 drives the expression of a different target gene set than TCF4. The pathways most affected by *LEF1* expression in colon cancer are EMT, interferon alpha, and TNFA signaling, demonstrating that LEF1 adds a lymphoid-like gene expression pattern to the colon cancer cell. This influences the growth and migration capabilities of the colon cancer cell as *LEF1* silencing, or loss, results in decreased growth rates and decreased rates of migration and invasion. It is also known that colon tumors interact with invading macrophages, in an attempt to inactivate the macrophages. The expression of immune regulatory genes, due to *LEF1* over-expression, may influence colon cancer-macrophage communication.

5.1.3 The TCF4-*LEF1* Interaction

After determining the influence of TCF4 and LEF1 on gene expression in the colon cancer cell line SW480, I sought to understand how TCF4 mediated repression of *LEF1*, which was decreased upon silencing of *TCF7L2*. ChIP-seq data demonstrated that TCF4 bound the *LEF1* promoter, suggesting a method for direct repression. There are two known families of transcriptional repressors which bind TCF4, the TLE family and the CtBP family. To determine whether the TLE family was mediating the repressive effect, I over-expressed *APC*, which facilitates the down-regulation of β-catenin, helping to convert TCF4 from an activator to a repressor. I used APC as opposed to an siRNA targeting β-catenin since the inherited familial adenomatous polyposis syndrome (FAP) is caused by mutation or loss of *APC*. I reasoned that use of APC and various APC fragments mimicking the various sizes of mutated *APC* could be used to degrade β-catenin, and down-regulate the WNT signaling pathway thereby increasing the likelihood that TCF4 functions as a repressor to further down-regulate *LEF1*. I would therefore be able to provide a plausible link between *LEF1* over-

expression and FAP. Results from plasmid transfections used to over-express *APC* or the various *APC* fragments, demonstrated that transfected SW480 cells had lower levels of total as well as nuclear β-catenin, and that these results were mirrored in a dual-luciferase assay. Transfected SW480 cells also grew more slowly than EV transfected cells, in accordance with decreased WNT signaling activity. Over-expression of *APC* as well as the *APC* 60, *APC* 70, and *APC* 80 fragments resulted in a ~25-50% reduction in the expression of *LEF1*, confirming my hypothesis. I then performed another set of *TCF7L2* silencing experiments with the *APC* wild-type colon cancer cell lines, HCT116 and RKO, as well as with two HCT116 clones and two RKO clones, which had biallelically mutated *APC*. These experiments demonstrated that while APC may influence *LEF1* expression in SW480, it is not a major regulator of *LEF1* expression in other cell lines, and other mechanisms may be at play.

I then explored the other family of transcriptional repressors, the CtBP family, and silenced the expression of the two CtBP genes in humans (*CtBP1* and *CtBP2*). Silencing of *CtBP1* resulted in a potent de-repression of *LEF1* expression, which did not influence the abundance of TCF4, eliminating any confounding effects due to varying TCF4 levels. These results were confirmed at the transcript level. I therefore conclude that TCF4 mediates direct repression of *LEF1* by binding with CtBP1, but not with CtBP2. In summation, I show that TCF4 is capable of simultaneously activating and repressing target gene expression in colon cancer, and that CtBP1 is required for this effect. TCF4 employs CtPB1 to maintain a TCF4-centric transcriptional program in colon cancer cells by repressing the expression of *LEF1* (Figure 43).

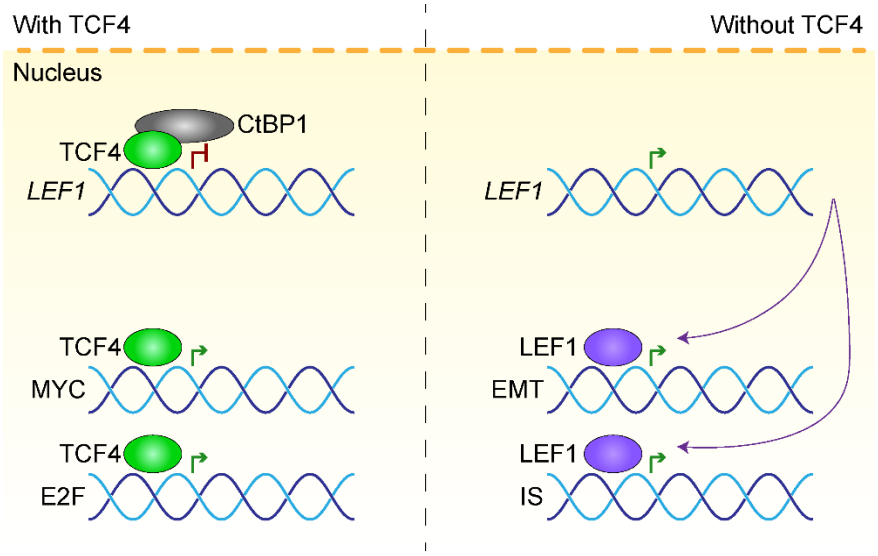

Figure 43 | Final diagram illustrating the findings of the dissertation.

5.1.4 Future Directions

I am interested in pursuing the influence of CtBP1 in other colon cancer cell lines to verify that the de-repression of *LEF1* is shared. Given that CtBP1 appears to repress WNT signaling transcription in an activated WNT background, it would be fascinating to further explore what factors mediate the ability of CtBP1 to repress WNT gene expression. For instance, is the strength of the repression predicted by the ratio of CtBP1 to TCF4? It appears that while TCF4 represses *LEF1*, it still activates the expression of other target genes. What determines the nature (activating or repressive) of TCF4 with its target gene, and is it possible to alter the target genes? For instance, can we repress *MYC*?

Another avenue to follow further explores the physical consequences of *TCF7L2* silencing and *LEF1* over-expression. I have generated a WNT reporter cell line which expresses a degron-tagged eGFP protein. The degron-tag shortens the half-life of eGFP so that changes in WNT signaling activity are more quickly realized at the protein level. The purpose of this cell line is to monitor WNT signaling activity while measuring other properties, such as invasion or migration. The SW480 reporter cell line also offers a useful test bed for quickly screening compounds to identify WNT signaling regulators without the need to transfect reporter plasmids and normalize for transfection efficiency. Given that loss of *TCF7L2* results in *LEF1* over-expression, which results in an increase in WNT signaling activity, these cells would prove useful in the identification of compounds which target both *TCF7L2* and *LEF1*. This is advantageous since compounds which only target TCF4, would not be able to address the *LEF1* feedback response, which may complicate treatment response.

Glossary

Benign: Non-invasive

Differentiated Cell: A cell with limited proliferative potential which fulfills a specific function in the body.

Epidemiology: The study of the incidence, distribution, and control of disease

Epithelium: A thin layer of cells which separates the body from the external environment.

Etiology: The cause, or set of causes, of a disease

Malignant: Invasive

Manifestation: A symptom or indicative marker of a disease

Mortality: Death

Mutation: An alteration in the DNA sequence

Stem Cell: A cell capable of nearly indefinite proliferative potential whose progeny can differentiate into the cell types necessary to perform the functions of a given tissue.

Tumor: A mass of cells exhibiting an abnormal growth pattern. A tumor may be comprised of benign or malignant cells, or a combination of the two.

Bibliography

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).
2. Cheng, X. *et al.* Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992-1997. *Cancer* **92**, 2547-54 (2001).
3. Murphy, G. *et al.* Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int J Cancer* **128**, 1668-75 (2011).
4. Gersten, O. & Wilmoth, J.R. The cancer transition in Japan since 1951. *Demographic Research* **7**, 271-306 (2002).
5. Fung, T. *et al.* Major dietary patterns and the risk of colorectal cancer in women. *Arch Intern Med* **163**, 309-14 (2003).
6. Chan, A.T. & Giovannucci, E.L. Primary prevention of colorectal cancer. *Gastroenterology* **138**, 2029-2043.e10 (2010).
7. Pisani, P., Parkin, D.M., Muñoz, N. & Ferlay, J. Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiol Biomarkers Prev* **6**, 387-400 (1997).
8. Parkin, D.M. *et al.* Part I: Cancer in Indigenous Africans--burden, distribution, and trends. *Lancet Oncol* **9**, 683-92 (2008).
9. Sitas, F. *et al.* Part II: Cancer in Indigenous Africans--causes and control. *Lancet Oncol* **9**, 786-95 (2008).
10. Omran, A.R. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem Fund Q* **49**, 509-38 (1971).
11. Sylla, B.S. & Wild, C.P. A million africans a year dying from cancer by 2030: what can cancer research and control offer to the continent? *Int J Cancer* **130**, 245-50 (2012).
12. Bray, F., Jemal, A., Grey, N., Ferlay, J. & Forman, D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol* **13**, 790-801 (2012).
13. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683-691 (2017).
14. Siegel, R.L. *et al.* Colorectal cancer statistics, 2020. *CA Cancer J Clin* **70**, 145-164 (2020).
15. Compton, C.C. Pathology report in colon cancer: what is prognostically important? *Dig Dis* **17**, 67-79 (1999).
16. Edwards, B.K. *et al.* Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **116**, 544-73 (2010).
17. Zauber, A.G. *et al.* Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* **366**, 687-96 (2012).
18. Bailey, C.E. *et al.* Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg* **150**, 17-22 (2015).

19. Siegel, R.L. *et al.* Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst* **109**(2017).
20. Kasi, P.M. *et al.* Rising Proportion of Young Individuals With Rectal and Colon Cancer. *Clin Colorectal Cancer* **18**, e87-e95 (2019).
21. Islami, F. *et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin* **68**, 31-54 (2018).
22. Wolf, A.M.D. *et al.* Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin* **68**, 250-281 (2018).
23. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**, 78-85 (2000).
24. Henrikson, N.B. *et al.* Family history and the natural history of colorectal cancer: systematic review. *Genet Med* **17**, 702-12 (2015).
25. Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M. & Wallace, M.B. Colorectal cancer. *Lancet* **394**, 1467-1480 (2019).
26. Syngal, S. *et al.* ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol* **110**, 223-62; quiz 263 (2015).
27. Ponz de Leon, M. & Di Gregorio, C. Pathology of colorectal cancer. *Dig Liver Dis* **33**, 372-88 (2001).
28. Marshman, E., Booth, C. & Potten, C.S. The intestinal epithelial stem cell. *Bioessays* **24**, 91-8 (2002).
29. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14**, e1002533 (2016).
30. Lipkin, M., Bell, B. & Sherlock, P. CELL PROLIFERATION KINETICS IN THE GASTROINTESTINAL TRACT OF MAN. I. CELL RENEWAL IN COLON AND RECTUM. *J Clin Invest* **42**, 767-76 (1963).
31. Schepers, A.G., Vries, R., van den Born, M., van de Wetering, M. & Clevers, H. Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. *EMBO J* **30**, 1104-9 (2011).
32. Peeters, T. & Vantrappen, G. The Paneth cell: a source of intestinal lysozyme. *Gut* **16**, 553-8 (1975).
33. Selsted, M.E., Miller, S.I., Henschen, A.H. & Ouellette, A.J. Enteric defensins: antibiotic peptide components of intestinal host defense. *J Cell Biol* **118**, 929-36 (1992).
34. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat Rev Gastroenterol Hepatol* **16**, 19-34 (2019).
35. Ayabe, T. *et al.* Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol* **1**, 113-8 (2000).
36. Radtke, F. & Clevers, H. Self-renewal and cancer of the gut: two sides of a coin. *Science* **307**, 1904-9 (2005).
37. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-7 (2007).
38. Escobar, M. *et al.* Intestinal epithelial stem cells do not protect their genome by asymmetric chromosome segregation. *Nat Commun* **2**, 258 (2011).

39. Lipkin, M. Proliferation and differentiation of gastrointestinal cells. *Physiol Rev* **53**, 891-915 (1973).
40. Lipkin, M. Phase 1 and phase 2 proliferative lesions of colonic epithelial cells in diseases leading to colonic cancer. *Cancer* **34**, suppl:878-88 (1974).
41. MESSIER, B. & LEBLOND, C.P. Cell proliferation and migration as revealed by radioautography after injection of thymidine-H3 into male rats and mice. *Am J Anat* **106**, 247-85 (1960).
42. Elson, C.O., Kagnoff, M.F., Fiocchi, C., Befus, A.D. & Targan, S. Intestinal immunity and inflammation: recent progress. *Gastroenterology* **91**, 746-68 (1986).
43. Fenoglio, C.M. & Lane, N. The anatomical precursor of colorectal carcinoma. *Cancer* **34**, suppl:819-23 (1974).
44. Takayama, T. *et al.* Aberrant crypt foci of the colon as precursors of adenoma and cancer. *N Engl J Med* **339**, 1277-84 (1998).
45. Cheng, L. & Lai, M.D. Aberrant crypt foci as microscopic precursors of colorectal cancer. *World J Gastroenterol* **9**, 2642-9 (2003).
46. Pretlow, T.P. *et al.* Aberrant crypts: putative preneoplastic foci in human colonic mucosa. *Cancer Res* **51**, 1564-7 (1991).
47. Roncucci, L., Medline, A. & Bruce, W.R. Classification of aberrant crypt foci and microadenomas in human colon. *Cancer Epidemiol Biomarkers Prev* **1**, 57-60 (1991).
48. Shpitz, B. *et al.* Aberrant crypt foci in human colons: distribution and histomorphologic characteristics. *Hum Pathol* **29**, 469-75 (1998).
49. Di Gregorio, C. *et al.* Histology of aberrant crypt foci in the human colon. *Histopathology* **30**, 328-34 (1997).
50. Roncucci, L. *et al.* Aberrant crypt foci in patients with colorectal cancer. *Br J Cancer* **77**, 2343-8 (1998).
51. Adler, D.G. *et al.* Endoscopic identification and quantification of aberrant crypt foci in the human colon. *Gastrointest Endosc* **56**, 657-62 (2002).
52. Soetikno, R.M. *et al.* Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults. *JAMA* **299**, 1027-35 (2008).
53. Lieberman, D., Moravec, M., Holub, J., Michaels, L. & Eisen, G. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. *Gastroenterology* **135**, 1100-5 (2008).
54. Aust, D.E., Baretton, G.B. & Pathology, M.o.t.W.G.G.-P.o.t.G.S.o. Serrated polyps of the colon and rectum (hyperplastic polyps, sessile serrated adenomas, traditional serrated adenomas, and mixed polyps)-proposal for diagnostic criteria. *Virchows Arch* **457**, 291-7 (2010).
55. MORSON, B.C. Precancerous lesions of the colon and rectum. Classification and controversial issues. *JAMA* **179**, 316-21 (1962).
56. Morson, B.C. The evolution of colorectal carcinoma. *Clin Radiol* **35**, 425-31 (1984).
57. Shinya, H. & Wolff, W.I. Morphology, anatomic distribution and cancer potential of colonic polyps. *Ann Surg* **190**, 679-83 (1979).
58. Muto, T., Bussey, H.J. & Morson, B.C. The evolution of cancer of the colon and rectum. *Cancer* **36**, 2251-70 (1975).
59. Strum, W.B. Colorectal Adenomas. *N Engl J Med* **375**, 389-90 (2016).

60. Eide, T.J. Natural history of adenomas. *World J Surg* **15**, 3-6 (1991).
61. Roncucci, L., Stamp, D., Medline, A., Cullen, J.B. & Bruce, W.R. Identification and quantification of aberrant crypt foci and microadenomas in the human colon. *Hum Pathol* **22**, 287-94 (1991).
62. Eide, T.J. Remnants of adenomas in colorectal carcinomas. *Cancer* **51**, 1866-72 (1983).
63. ENTERLINE, H.T., EVANS, G.W., MERCUDO-LUGO, R., MILLER, L. & FITTS, W.T. Malignant potential of adenomas of colon and rectum. *JAMA* **179**, 322-30 (1962).
64. Jen, J. *et al.* Molecular determinants of dysplasia in colorectal lesions. *Cancer Res* **54**, 5523-6 (1994).
65. Cardoso, J., Boer, J., Morreau, H. & Fodde, R. Expression and genomic profiling of colorectal cancer. *Biochim Biophys Acta* **1775**, 103-37 (2007).
66. Fleming, M., Ravula, S., Tatishchev, S.F. & Wang, H.L. Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol* **3**, 153-73 (2012).
67. Compton, C.C. *et al.* Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* **124**, 979-94 (2000).
68. DUKES, C.E. & BUSSEY, H.J. The spread of rectal cancer and its effect on prognosis. *Br J Cancer* **12**, 309-20 (1958).
69. D'Eredita, G. *et al.* A survival regression analysis of prognostic factors in colorectal cancer. *Aust N Z J Surg* **66**, 445-51 (1996).
70. Edge, S.B. & American Joint Committee on Cancer. *AJCC cancer staging manual*, xiv, 648 p. (Springer, New York, 2010).
71. Ponz de Leon, M. *et al.* Clinical and pathologic prognostic indicators in colorectal cancer. A population-based study. *Cancer* **69**, 626-35 (1992).
72. Roncucci, L. *et al.* Survival for colon and rectal cancer in a population-based cancer registry. *Eur J Cancer* **32A**, 295-302 (1996).
73. Sasaki, O., Atkin, W.S. & Jass, J.R. Mucinous carcinoma of the rectum. *Histopathology* **11**, 259-72 (1987).
74. Symonds, D.A. & Vickery, A.L. Mucinous carcinoma of the colon and rectum. *Cancer* **37**, 1891-1900 (1976).
75. Alexander, J. *et al.* Histopathological identification of colon cancer with microsatellite instability. *Am J Pathol* **158**, 527-35 (2001).
76. Leopoldo, S. *et al.* Two subtypes of mucinous adenocarcinoma of the colorectum: clinicopathological and genetic features. *Ann Surg Oncol* **15**, 1429-39 (2008).
77. Chen, J.S. *et al.* Clinical outcome of signet ring cell carcinoma and mucinous adenocarcinoma of the colon. *Chang Gung Med J* **33**, 51-7 (2010).
78. LAUFMAN, H. & SAPHIR, O. Primary linitis plastica type of carcinoma of the colon. *AMA Arch Surg* **62**, 79-91 (1951).
79. Mizushima, T. *et al.* Primary colorectal signet-ring cell carcinoma: clinicopathological features and postoperative survival. *Surg Today* **40**, 234-8 (2010).
80. Pande, R. *et al.* Significance of signet-ring cells in patients with colorectal cancer. *Dis Colon Rectum* **51**, 50-5 (2008).
81. Thirunavukarasu, P. *et al.* Medullary carcinoma of the large intestine: a population based analysis. *Int J Oncol* **37**, 901-7 (2010).

82. Lanza, G., Gafà, R., Matteuzzi, M. & Santini, A. Medullary-type poorly differentiated adenocarcinoma of the large bowel: a distinct clinicopathologic entity characterized by microsatellite instability and improved survival. *J Clin Oncol* **17**, 2429-38 (1999).
83. Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-70 (1996).
84. Bocker, T. *et al.* Genomic instability in colorectal carcinomas: comparison of different evaluation methods and their biological significance. *J Pathol* **179**, 15-9 (1996).
85. Powell, S.M. *et al.* APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235-7 (1992).
86. Kinzler, K.W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763 (1997).
87. Miyoshi, Y. *et al.* Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc Natl Acad Sci U S A* **89**, 4452-6 (1992).
88. Smith, K.J. *et al.* The APC gene product in normal and tumor cells. *Proc Natl Acad Sci U S A* **90**, 2846-50 (1993).
89. Powell, S.M. *et al.* Molecular diagnosis of familial adenomatous polyposis. *N Engl J Med* **329**, 1982-7 (1993).
90. Smith, A.J. *et al.* Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. *Cancer Res* **54**, 5527-30 (1994).
91. Ried, T. *et al.* Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* **15**, 234-45 (1996).
92. Ried, T. Homage to Theodor Boveri (1862-1915): Boveri's theory of cancer as a disease of the chromosomes, and the landscape of genomic imbalances in human carcinomas. *Environ Mol Mutagen* **50**, 593-601 (2009).
93. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-32 (1988).
94. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-58 (2013).
95. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* **105**, 4283-8 (2008).
96. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-67 (1990).
97. Baker, S.J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217-21 (1989).
98. Yamashita, N., Minamoto, T., Ochiai, A., Onda, M. & Esumi, H. Frequent and characteristic K-ras activation and absence of p53 protein accumulation in aberrant crypt foci of the colon. *Gastroenterology* **108**, 434-40 (1995).
99. Losi, L., Roncucci, L., di Gregorio, C., de Leon, M.P. & Benhattar, J. K-ras and p53 mutations in human colorectal aberrant crypt foci. *J Pathol* **178**, 259-63 (1996).
100. Pretlow, T.P., Brasitus, T.A., Fulton, N.C., Cheyer, C. & Kaplan, E.L. K-ras mutations in putative preneoplastic lesions in human colon. *J Natl Cancer Inst* **85**, 2004-7 (1993).

101. Jass, J.R. *et al.* Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *J Clin Pathol* **52**, 455-60 (1999).
102. Yang, S., Farraye, F.A., Mack, C., Posnik, O. & O'Brien, M.J. BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *Am J Surg Pathol* **28**, 1452-9 (2004).
103. Kambara, T. *et al.* BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* **53**, 1137-44 (2004).
104. Bettington, M. *et al.* The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* **62**, 367-86 (2013).
105. Snover, D.C. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol* **42**, 1-10 (2011).
106. Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156-60 (2005).
107. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999 (2017).
108. Miller, J.R. The Wnts. *Genome Biol* **3**, REVIEWS3001 (2002).
109. McMahon, A.P., Joyner, A.L., Bradley, A. & McMahon, J.A. The midbrain-hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell* **69**, 581-95 (1992).
110. Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448-52 (2003).
111. Takada, R. *et al.* Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev Cell* **11**, 791-801 (2006).
112. Rios-Esteves, J., Haugen, B. & Resh, M.D. Identification of key residues and regions important for porcupine-mediated Wnt acylation. *J Biol Chem* **289**, 17009-19 (2014).
113. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* **125**, 523-33 (2006).
114. Bänziger, C. *et al.* Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125**, 509-22 (2006).
115. Port, F. *et al.* Wingless secretion promotes and requires retromer-dependent cycling of Wntless. *Nat Cell Biol* **10**, 178-85 (2008).
116. Jue, S.F., Bradley, R.S., Rudnicki, J.A., Varmus, H.E. & Brown, A.M. The mouse Wnt-1 gene can act via a paracrine mechanism in transformation of mammary epithelial cells. *Mol Cell Biol* **12**, 321-8 (1992).
117. Korkut, C. *et al.* Trans-synaptic transmission of vesicular Wnt signals through Evi/Wntless. *Cell* **139**, 393-404 (2009).
118. Gross, J.C., Chaudhary, V., Bartscherer, K. & Boutros, M. Active Wnt proteins are secreted on exosomes. *Nat Cell Biol* **14**, 1036-45 (2012).
119. Janda, C.Y., Waghray, D., Levin, A.M., Thomas, C. & Garcia, K.C. Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59-64 (2012).
120. Janda, C.Y. *et al.* Surrogate Wnt agonists that phenocopy canonical Wnt and β-catenin signalling. *Nature* **545**, 234-237 (2017).
121. Bhanot, P. *et al.* A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382**, 225-30 (1996).

122. Hsieh, J.C., Rattner, A., Smallwood, P.M. & Nathans, J. Biochemical characterization of Wnt-frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc Natl Acad Sci U S A* **96**, 3546-51 (1999).
123. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530-5 (2000).
124. Tamai, K. *et al.* A mechanism for Wnt coreceptor activation. *Mol Cell* **13**, 149-56 (2004).
125. Mao, J. *et al.* Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol Cell* **7**, 801-9 (2001).
126. Cliffe, A., Hamada, F. & Bienz, M. A role of Dishevelled in relocating Axin to the plasma membrane during wingless signaling. *Curr Biol* **13**, 960-6 (2003).
127. Schwarz-Romond, T., Merrifield, C., Nichols, B.J. & Bienz, M. The Wnt signalling effector Dishevelled forms dynamic protein assemblies rather than stable associations with cytoplasmic vesicles. *J Cell Sci* **118**, 5269-77 (2005).
128. Li, V.S. *et al.* Wnt signaling through inhibition of β-catenin degradation in an intact Axin1 complex. *Cell* **149**, 1245-56 (2012).
129. Molenaar, M. *et al.* XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell* **86**, 391-9 (1996).
130. Behrens, J. *et al.* Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature* **382**, 638-42 (1996).
131. van de Wetering, M., Oosterwegel, M., Dooijes, D. & Clevers, H. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J* **10**, 123-32 (1991).
132. Waterman, M.L., Fischer, W.H. & Jones, K.A. A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev* **5**, 656-69 (1991).
133. Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev* **5**, 880-94 (1991).
134. Oosterwegel, M. *et al.* Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J Exp Med* **173**, 1133-42 (1991).
135. Castrop, J., van Norren, K. & Clevers, H. A gene family of HMG-box transcription factors with homology to TCF-1. *Nucleic Acids Res* **20**, 611 (1992).
136. Cadigan, K.M. & Waterman, M.L. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol* **4**(2012).
137. van de Wetering, M., Oosterwegel, M., van Norren, K. & Clevers, H. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J* **12**, 3847-54 (1993).
138. Giese, K., Amsterdam, A. & Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev* **5**, 2567-78 (1991).
139. Giese, K., Cox, J. & Grosschedl, R. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69**, 185-95 (1992).
140. Love, J.J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-5 (1995).

141. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789-99 (1997).
142. Cavallo, R.A. *et al.* Drosophila Tcf and Groucho interact to repress Wingless signalling activity. *Nature* **395**, 604-8 (1998).
143. Calvo, D. *et al.* A POP-1 repressor complex restricts inappropriate cell type-specific gene transcription during Caenorhabditis elegans embryogenesis. *EMBO J* **20**, 7197-208 (2001).
144. Daniels, D.L. & Weis, W.I. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol* **12**, 364-71 (2005).
145. Arce, L., Pate, K.T. & Waterman, M.L. Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. *BMC Cancer* **9**, 159 (2009).
146. Brannon, M., Brown, J.D., Bates, R., Kimelman, D. & Moon, R.T. XCtBP is a XTcf-3 co-repressor with roles throughout Xenopus development. *Development* **126**, 3159-70 (1999).
147. Valenta, T., Lukas, J. & Korinek, V. HMG box transcription factor TCF-4's interaction with CtBP1 controls the expression of the Wnt target Axin2/Conductin in human embryonic kidney cells. *Nucleic Acids Res* **31**, 2369-80 (2003).
148. Atcha, F.A. *et al.* A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol Cell Biol* **27**, 8352-63 (2007).
149. Chang, M.V., Chang, J.L., Gangopadhyay, A., Shearer, A. & Cadigan, K.M. Activation of wingless targets requires bipartite recognition of DNA by TCF. *Curr Biol* **18**, 1877-81 (2008).
150. Verbeek, S. *et al.* An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**, 70-4 (1995).
151. Roose, J. *et al.* Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science* **285**, 1923-6 (1999).
152. Merrill, B.J. *et al.* Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development* **131**, 263-74 (2004).
153. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171-83 (2006).
154. Hrckulak, D., Kolar, M., Strnad, H. & Korinek, V. TCF/LEF Transcription Factors: An Update from the Internet Resources. *Cancers (Basel)* **8**(2016).
155. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* **19**, 379-83 (1998).
156. van Es, J.H. *et al.* A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. *Mol Cell Biol* **32**, 1918-27 (2012).
157. Korinek, V. *et al.* Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**, 1784-7 (1997).
158. Morin, P.J. *et al.* Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-90 (1997).
159. van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev* **8**, 2691-703 (1994).
160. Liu, C. *et al.* Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**, 837-47 (2002).

161. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. beta-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J* **16**, 3797-804 (1997).
162. Kitagawa, M. *et al.* An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of beta-catenin. *EMBO J* **18**, 2401-10 (1999).
163. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer* **3**, 11-22 (2003).
164. Bos, J.L. ras oncogenes in human cancer: a review. *Cancer Res* **49**, 4682-9 (1989).
165. Quilliam, L.A., Khosravi-Far, R., Huff, S.Y. & Der, C.J. Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. *Bioessays* **17**, 395-404 (1995).
166. Boguski, M.S. & McCormick, F. Proteins regulating Ras and its relatives. *Nature* **366**, 643-54 (1993).
167. Moodie, S.A. & Wolfman, A. The 3Rs of life: Ras, Raf and growth regulation. *Trends Genet* **10**, 44-8 (1994).
168. Campbell, S.L., Khosravi-Far, R., Rossman, K.L., Clark, G.J. & Der, C.J. Increasing complexity of Ras signaling. *Oncogene* **17**, 1395-413 (1998).
169. Crews, C.M. & Erikson, R.L. Extracellular signals and reversible protein phosphorylation: what to Mek of it all. *Cell* **74**, 215-7 (1993).
170. Crews, C.M., Alessandrini, A. & Erikson, R.L. Erks: their fifteen minutes has arrived. *Cell Growth Differ* **3**, 135-42 (1992).
171. Marais, R., Wynne, J. & Treisman, R. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell* **73**, 381-93 (1993).
172. Massagué, J. TGFβ signalling in context. *Nat Rev Mol Cell Biol* **13**, 616-30 (2012).
173. Shi, Y. & Massagué, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).
174. Wrana, J.L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF-beta receptor. *Nature* **370**, 341-7 (1994).
175. Massagué, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev* **19**, 2783-810 (2005).
176. Mullen, A.C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* **147**, 565-76 (2011).
177. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
178. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-9 (2016).
179. Brown, M.A. *et al.* Not any good, this one. TCF7L2 silencing results in altered gene expression patterns accompanied by local genomic reorganization. *Neoplasia* **23**, 257-269 (2021).